

**Applying Stochastic Programming
Models in Financial Risk
Management**

Xi Yang

Doctor of Philosophy
University of Edinburgh
2009

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(Xi Yang)

Abstract

This research studies two modelling techniques that help seek optimal strategies in financial risk management. Both are based on the stochastic programming methodology. The first technique is concerned with market risk management in portfolio selection problems; the second technique contributes to operational risk management by optimally allocating workforce from a managerial perspective.

The first model involves multiperiod decisions (portfolio rebalancing) for an asset and liability management problem and deals with the usual uncertainty of investment returns and future liabilities. Therefore it is well-suited to a stochastic programming approach. A stochastic dominance concept is applied to control the risk of underfunding. A small numerical example and a backtest are provided to demonstrate advantages of this new model which includes stochastic dominance constraints over the basic model.

Adding stochastic dominance constraints comes with a price: it complicates the structure of the underlying stochastic program. Indeed, new constraints create a link between variables associated with different scenarios of the same time stage. This destroys the usual tree-structure of the constraint matrix in the stochastic program and prevents the application of standard stochastic programming approaches such as (nested) Benders decomposition and progressive hedging. A structure-exploiting interior point method is applied to this problem. Computational results on medium scale problems with sizes reaching about one million variables demonstrate the efficiency of the specialised solution technique.

The second model deals with operational risk from human origin. Unlike market risk that can be handled in a financial manner (e.g. insurances, savings, derivatives), the treatment of operational risks calls for a “managerial approach”. Consequently, we propose a new way of dealing with operational risk, which relies on the well known Aggregate Planning Model. To illustrate this idea, we have adapted this model to the case of a back office of a bank specialising in the trading of derivative products. Our contribution corresponds to several improvements applied to stochastic programming modelling. First, the basic model is transformed into a multistage stochastic program in order to take into account the randomness associated with the volume of transaction demand and with the capacity of work provided by qualified and non-qualified employees over the planning horizon. Second, as advocated by Basel II, we calculate the probability distribution based on a Bayesian Network to circumvent the difficulty of obtaining data which characterises uncertainty in operations. Third, we go a step further by relaxing the traditional assumption in stochastic programming that imposes a strict independence between the decision variables and the random elements. Comparative results show that in general these improved stochastic programming models tend to allocate more human expertise in order to hedge operational risks. The dual solutions of the stochastic programs are exploited to detect periods and nodes that are at risk in terms of expertise availability.

Acknowledgements

I would like to take this opportunity to show my gratitude to Prof. Jacek Gondzio for supervising me on my research and navigating me whenever I lost my direction along the way. He shared his enthusiasm on academia and encouraged me to enjoy the academic work out of frustrations.

Andreas Grothey contributed a lot of his valuable time and effort on this project. He has had the patience of introducing the interior point solver OOPS to me and assisted me at the deep end of working on OOPS.

I would also like to thank Emmanuel Fragnière for proposing the interesting and challenging topic on operational risk management and his brilliant work on co-writing a paper on operational risk management.

I also acquired many research ideas from Prof. Ken I M McKinnon, whose constructive comments helped me justify my work.

Marco Colombo always gave me solutions when I was faced with problems and turned to him for help. Andrew Thompson's effort on proofreading my thesis is also much appreciated.

My family gave me great support and had always been encouraging me throughout my PhD life. Their pride in me inspires me to do my work to the best possible standard. I also received a good deal of support from Ge Liu who is close to me and went through the hard time with me.

Contents

Abstract	3
1 Introduction	7
1.1 Financial Risk Management	7
1.2 Market Risk	8
1.2.1 Utility	8
1.2.2 Mean-Risk Model	10
1.2.3 Coherent Risk Measure	12
1.2.4 Asset-Liability Management	16
1.3 Operational Risk	16
1.3.1 Definition of Operational Risk	17
1.3.2 Measuring and Managing Operational Risk	17
1.3.3 Workforce Planning Problem	19
1.4 Stochastic Programming	20
1.5 Outline of the Thesis	23
2 Some Modelling and Solution Techniques of Stochastic Programming	24
2.1 Induction of Stochastic Programming	24
2.2 Recourse Problems	25
2.2.1 Linear Recourse Programming	26
2.2.2 Multiple-Stage Recourse Problems	28
2.3 Scenario Generation	32
2.4 Solving Stochastic Linear Programming	32
2.4.1 Duality Theory and Shadow Prices	32
2.4.2 Benders Decomposition	34
2.4.3 Interior Point Method and OOPS	37
2.5 Exogenous and Endogenous Problems	42
2.6 Application Areas	44
3 Measuring and Managing Market Risk by Stochastic Dominance	46
3.1 Asset-Liability Management	46
3.1.1 Literature Review of Asset-Liability Management Modelling . . .	46
3.1.2 Multi-Stage ALM Modelling	49

3.2	Stochastic Dominance	52
3.2.1	Definitions of Stochastic Dominance	52
3.2.2	Literature Review of Stochastic Dominance	54
3.2.3	Monotonicity of CVaR with SSD	56
3.2.4	Linear Formulation of SSD	58
3.2.5	Interval Second-order Stochastic Dominance and Chance Con- straints	59
3.3	Multi-Stage ALM Model with SSD and Relaxed ISSD Constraints . . .	65
3.4	Numerical Results	67
3.4.1	A Model Example	67
3.4.2	Backtesting	70
3.4.3	Numerical Efficiency	72
3.4.4	Parallel Solution	77
4	Qualified Workforce Capacity Planning in Operational Risk Manage- ment	79
4.1	Workforce Planning Problem	79
4.2	Aggregate Planning Model	81
4.2.1	Generic Aggregate Planning Model	81
4.2.2	Deterministic Work Force Planning Model	83
4.3	Multi-stage Stochastic Aggregate Planning Model of Workforce Planning Problem	85
4.3.1	Random Demand Parameters	85
4.3.2	Random Capacity Parameters	86
4.4	Revising Operation Efficiency Probability Distributions	87
4.4.1	Bayesian Network	88
4.4.2	Probability Distribution Revision	94
4.5	Random Parameters Dependent on Decisions	96
4.6	Implementations of the Models	100
4.7	The Pricing of Operational Risk	102
4.7.1	The Sensitivity Analysis of Nonlinear Programming	103
4.7.2	Exploiting the Shadow Pricing Approach	105
5	Conclusions and Future Work	107
5.1	Research Outcomes	107
5.2	Research Prospects	109
	Bibliography	110
A	Parameters and results of Model (4.12) in Chapter 4	120

Chapter 1

Introduction

This research is based on the methodology of stochastic programming, applied to measure and manage the financial risk. In this chapter we introduce the problems considered and the main and basic methodology used in this project. An outline of this thesis is provided at the end.

1.1 Financial Risk Management

On 11th of December 2008, the former chairman of the NASDAQ stock exchange and the Wall Street firm Bernard L. Madoff Investment Securities LLC, Bernard Lawrence Madoff was arrested by the FBI on the allegation of a \$50bn fraud, the largest fraud in history. This fraud operated as a Ponzi scheme, which pays investors with their own money or the money collected from new investors. HSBC had potential exposure of \$1bn; the exposure for Royal Bank of Scotland was £400m; about \$468m for France's BNP Paribas; Spain's Banco Santander said it had a direct exposure of 17m Euros; etc.

Due to the development of financial services and financial products, financial business is becoming more and more complicated, resulting in greater risks. In addition, the financial industry is getting more involved in the economy, both nationally and globally. The financial market has an influence on economy to a certain extent, e.g. the gloom of the financial market could lead to an economic recession. Subprime mortgages, mono-line insurers, collateralised debt obligations, the collapse of Lehman Brothers, bail-outs for everyone from AIG to the Royal Bank of Scotland and one certain Bernard Madoff; all of these warn the world that better financial risk control is desperately needed.

The Basel II framework [88], produced by the Basel Committee on Banking Supervision which is located at BIS (Bank for International Settlement), helps financial institutions to identify risk and manage it by regulatory capital requirements. It has been widely applied in financial risk management. As the world's oldest international financial organisation, BIS fosters international monetary and financial corporations as a bank of central banks, with 55 central bank members. The Basel Committee on Banking Supervision, one of the five standing committees supporting central banks

and authorities in charge of financial stability by providing background analysis and policy recommendations, helps enhance understanding of key supervisory issues and improve the quality of banking supervision worldwide, while also developing guidelines and supervisory standards in desirable areas.

Basel II [88] introduced three pillars for measuring and managing financial risk: minimum capital requirement, supervisory review process and market discipline. Minimum capital requirement guides the institutions on the calculation of capital that should be set aside to cover the risk; it is followed by the supervisory review process to enhance the responsibility of bank management to stick to the first pillar; then market discipline warrants the introduction of disclosure requirements for banks using the framework. The three main categories of financial risk under the analysis of Basel II are market risk, credit risk and operational risk. Market risk is mainly from the fluctuations of asset prices in the financial market. Credit risk corresponds to the default of counterparties or business institutions involved in the financial market, i.e. failure to comply with their obligation to service debt. Unlike market risk and credit risk, it is difficult to give operational risk a clear-cut definition. An easy way to understand operational risk is the risk involved in operations. More comprehensive definitions of operational risk are given in Section 1.3. In this work, we are concerned with market risk and operational risk.

1.2 Market Risk

When trading products or managing assets in financial markets, one has to consider the risk due to market volatility, i.e. market risk. When pursuing an investment program, different people have different attitudes to the profit they could earn and the risk they consider acceptable. Hence, in modern portfolio theory, utility functions are applied to measure the performance of the portfolio and the satisfaction of investors, who will choose to maximise the expected utility, see [81]. One way to describe and optimise utility is the mean-risk model, which originated in the late 1950s. Since then, various risk measures have been used with variance and Value-at-Risk as two of the most popular ones.

1.2.1 Utility

To know the purpose is the first task in constructing a portfolio to purchase. Generally, the simple objective is to maximise our wealth with acceptable risk. Are people's attitudes the same toward wealth and risk? Even for the same portfolio? Suppose there are two portfolios, i.e. one has returns of £100 and £0 with equal probabilities, the other has return of £50 for sure. The expected returns of these two portfolios are the same. A person who has only £10 in total will be more inclined to choose the latter one. Another with property of over £10,000 is likely to prefer to gamble on the first

one. This is because these two portfolios each present different utilities to these two people.

The Swiss mathematician Daniel Bernoulli first proposed the notion of utility in a paper named “Exposition of a New Theory on the Measure of Risk” [11], which was published in 1738. He pointed out that the portfolio value must not be determined only by its price, but also by the utility it yields, where utility is a measure of a degree of satisfaction.

“The price of the item is dependent only on the thing itself and is equal for everyone; the utility, however is dependent on the particular circumstances of the person making the estimate.”

Two significant properties of utility were also discussed:

- any increase in wealth, no matter how insignificant, will always result in an increase in utility;
- increase in utility is inversely proportional to the quantity of wealth already possessed.

In portfolio selection, utility functions are constructed according to investors’ attitude to wealth and risk. Generally speaking, investors prefer more wealth. For each portfolio pattern, a utility function associates a number to it. Figure 1.1 is an exam-

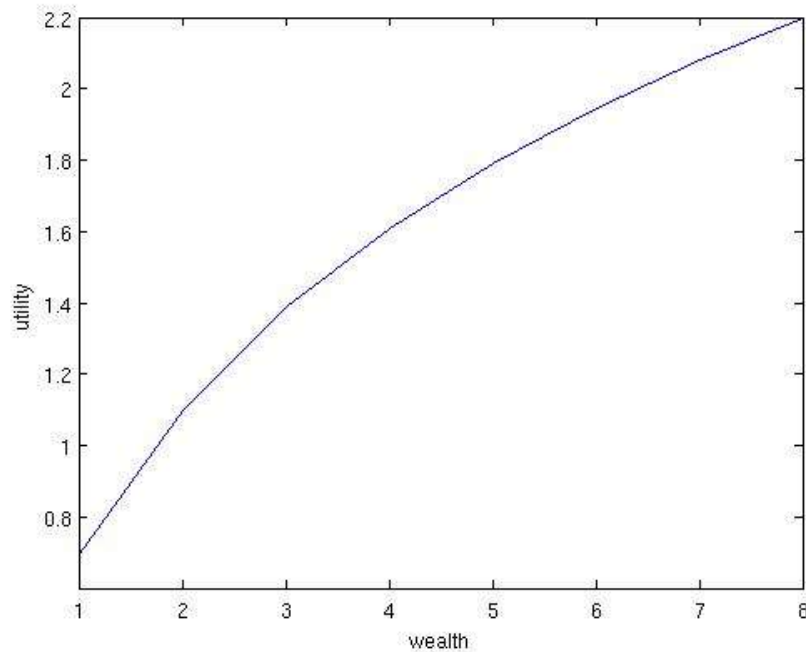


Figure 1.1: An example of utility function

ple of a risk-averse utility function, which is a log function. See more in [84]. The horizontal axis represents wealth, while the vertical axis represents utility. The curve measures the relationship between wealth and utility. The positive slope indicates that investors prefer more wealth to less wealth. The fact that the steepness of the slope decreases as wealth increases implies that investors derive less and less satisfaction with each increment unit of wealth. Both reflect two properties of utility. Such monotonic increasing and concave functions are applicable in most cases. When utility functions are used in portfolio selection where the return of the portfolio is uncertain, expected utility maximisation serves as good judgement.

While it is not easy to find out the precise utility function for each investor, we can also order utilities of portfolios by preference instead. Stochastic dominance manages to rank portfolios consistent with general utility functions. In Chapter 3, we will show how stochastic dominance can be used to control and manage market risk so as to construct an optimal portfolio strategy.

1.2.2 Mean-Risk Model

In portfolio construction, we are concerned with two aspects: return and risk. Since the return is uncertain, we usually consider its expected value. Risk represents the variability of the portfolio value. An optimal portfolio should attain as large an expected value as possible and this inevitably incurs considerable risk: therefore, we need to trade off these two factors against each other. Mean-risk model, based on the stochastic programming methodology which allows uncertainty to be taken into account, provides a general framework to construct portfolio considering investors' utility functions. The efficient frontier described later can illustrate investors' attitude to the return and risk of the portfolio.

Suppose a financial institution plans to invest in assets from a set $I = \{1, \dots, m\}$, with x_i denoting the units invested in asset i . The return r_i of asset i has a known probability distribution and the total return of the portfolio is R . We make the strong assumption that the probability distribution can be deduced (approximated) from historical data. Then we can calculate the expected return of the portfolio:

$$E[R] = \sum_{i \in I} E[x_i r_i] = \sum_{i \in I} x_i E[r_i].$$

Considering a risk function $\phi(x)$ measuring the risk incurred by decision $x \in X \subseteq R^m$ where X is some feasible set of x , a general portfolio selection problem can be formulated in one of the following three ways:

$$\min_x \quad -E[r'x] + \gamma\phi(x), \quad x \in X, \quad (1.1)$$

$$\min_x \quad \phi(x), \quad E[r'x] \geq R, \quad x \in X, \quad (1.2)$$

$$\min_x \quad -E[r'x], \quad \phi(x) \leq \beta, \quad x \in X, \quad (1.3)$$

where γ is the risk aversion parameter, R and β is the return and risk threshold, respectively. r' is the transpose of r . Suppose that the constraints $E[r'x] \geq R$, $\phi(x) \leq \beta$ admit strictly feasible points. It is proved in [76] that given a convex set X and a convex risk measure function $\phi(x)$, these three problems are equivalent in the sense that they can generate the same efficient frontier. The best-known example of formulation (1.1) is the Markowitz mean-variance multi-objective model [86], which considers both return and risk in the objective. For this work, Harry Markowitz received the 1990 Nobel Prize in Economics. In formulation (1.2) risk is minimised with acceptable returns, while in formulation (1.3), the return is maximised subject to risk being kept at an acceptable level. The constraint in (1.3) defines the feasible set with feasible risk so that in the objective the decision-maker can focus on maximising the return.

Several constraints should be considered in practice. The first is the budget constraint, i.e. how much wealth should be invested initially. The second is the value constraint in each asset: the minimum and maximum value that can be invested in a particular asset, as well as a bound on the position change. In addition, the transaction cost may be incorporated if multiple stages are considered. In practice, there may also be some legal and policy constraints.

By varying the risk tolerance β or coefficient of risk term γ in the model, we can generate an efficient frontier of portfolios, i.e. the portfolios on the curve are preferable to those under the curve and the points above the efficient frontier are infeasible. For example, in Figure 1.2 the curve given by the border of the dark area is the efficient frontier. The whole dark area provides all the feasible solutions and the area above the curve is infeasible. Such a frontier is efficient because compared to other portfolios, these strategies can either produce highest return with the same risk or bear lowest risk to achieve the same return.

Variance, as one of the key statistical parameters, has been used to measure market risk in mean-risk model since Markowitz mean-variance model. It is easy to implement and simple to understand, i.e. reflecting the mean distance of random values to the expected value. Moreover, because variance has been well analysed and is familiar to people, this facilitates its use as a risk measure. Its application, under two specific cases, i.e. quadratic utility function and symmetric distributions of returns (e.g. normal distributions, lognormal distributions), can provide investors with a proper assessment of portfolios. Without these assumptions, however, variance has several drawbacks and could even lead to inferior investment strategies. A straightforward explanation is that variance considers extremely high and extremely low returns equally undesirable. Besides downside risk, variance also takes upside variability as risk. The analysis of pros and cons of variance can be found in Chapter 9 in [86]. From the risk measure's perspective, variance is not coherent. The following section will give the definition of coherence of risk measure with examples of such measures.

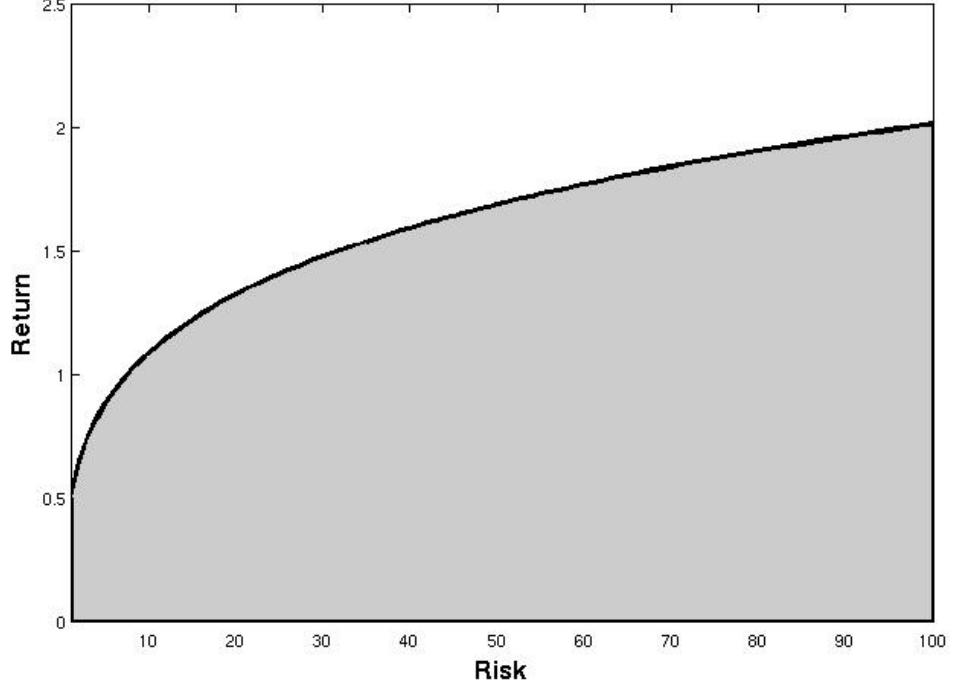


Figure 1.2: An example of efficient frontier.

1.2.3 Coherent Risk Measure

Following [6], a coherent risk measure is defined as:

Definition 1. A risk measure $\phi(\xi)$, where $\xi \in \Omega$ is a random variable as the future value of a portfolio with probability distribution function F_ξ , is called coherent if it satisfies the following four conditions:

1. *Translation invariance:* for all $\xi \in \Omega$ and all real numbers α , we have $\phi(\xi + \alpha) = \phi(\xi) - \alpha$;
2. *Subadditivity:* for all ξ_1 and $\xi_2 \in \Omega$, $\phi(\xi_1 + \xi_2) \leq \phi(\xi_1) + \phi(\xi_2)$.
3. *Positive homogeneity:* for all $\lambda \geq 0$ and all $\xi \in \Omega$, $\phi(\lambda\xi) = \lambda\phi(\xi)$.
4. *Monotonicity:* for all ξ_1 and $\xi_2 \in \Omega$ with $\xi_1 \leq \xi_2$, we have $\phi(\xi_2) \leq \phi(\xi_1)$.

Translation invariance means that by adding a sure amount α to the portfolio, the risk will be reduced by α because the future value of the portfolio will increase by α . Subadditivity simply illustrates the diversification of the portfolio. Although subadditivity implies that $\rho(nX) \leq n\rho(X)$, multiplying the same position cannot lead to diversification and the positive homogeneity holds. Monotonicity is natural. A representation of coherence is exploited in [112] using convex analysis from a topology perspective.

Remark 1.1. *It is not difficult to deduce from subadditivity and positive homogeneity that: $\rho(\alpha\xi_1 + (1 - \alpha)\xi_2) \leq \alpha\rho(\xi_1) + (1 - \alpha)\rho(\xi_2)$, for $0 \leq \alpha \leq 1$.*

Several risk measures have been proved to be coherent, including conditional Value-at-Risk and semideviation. In the following part, we will discuss Value-at-Risk and conditional Value-at-Risk, which have been attracting more and more attention from financial industry.

Value-at-Risk [4, 76, 77, 79, 104, 109, 122, 123] is advised by the Basel committee and has been accepted and used in a lot of financial institutions. In financial applications, the percentile of the loss is called Value-at-Risk (VaR). It describes the maximum loss (or other measure of performance) with a specified confidence level. Let $f(x, \xi)$ be the loss (or other measure of performance) associated with the decision vector x , to be chosen from a certain subset X of \mathbb{R}^n , and the random vector ξ in \mathbb{R}^m . X can be interpreted as representing a portfolio, such as the investment units on certain portfolios or assets. The vector ξ stands for the uncertainties, e.g. market variables, that can affect the loss. For each x , the loss $f(x, \xi)$ is a random variable having a distribution in \mathbb{R} induced by that of ξ . The underlying probability of ξ in \mathbb{R}^m is assumed for convenience to have a density, which we denote $p(\xi)$. The probability of $f(x, \xi)$ not exceeding a threshold α is then given by

$$\Phi(x, \alpha) = \int_{f(x, \xi) \leq \alpha} p(\xi) d\xi.$$

$\Phi(x, \alpha)$ is the cumulative distribution function for the performance $f(x, \xi)$ depending on x . It is nondecreasing and right continuous with α .

β -VaR i.e. β -quantile for the performance random variable associated with x and any specified probability level $\beta \in (0, 1)$, is defined as:

$$\alpha_\beta(x) = \min\{\alpha \in \mathbb{R} : \Phi(x, \alpha) \geq \beta\}.$$

Since $\Phi(x, \alpha)$ is right continuous and nondecreasing, β -VaR appears as the left endpoint of the nonempty interval consisting of the values α such that $F(x, \alpha) = \beta$, see [4]. Hence, we can see that β -VaR gives the lowest loss amount α of the portfolio and the loss of the portfolio will not exceed this amount with probability β , i.e.

$$\Phi(x, \alpha_\beta(x)) \geq \beta. \tag{1.4}$$

β -CVaR defined as:

$$\phi_\beta(x) = (1 - \beta)^{-1} \int_{f(x, \xi) \geq \alpha_\beta(x)} f(x, \xi) p(\xi) d\xi,$$

is actually the conditional expectation of the performance $f(x, \xi)$ which exceeds the β -VaR. We can see that the probability of $f(x, \xi) \geq \alpha_\beta(x)$ is just $1 - \beta$ in the integration,

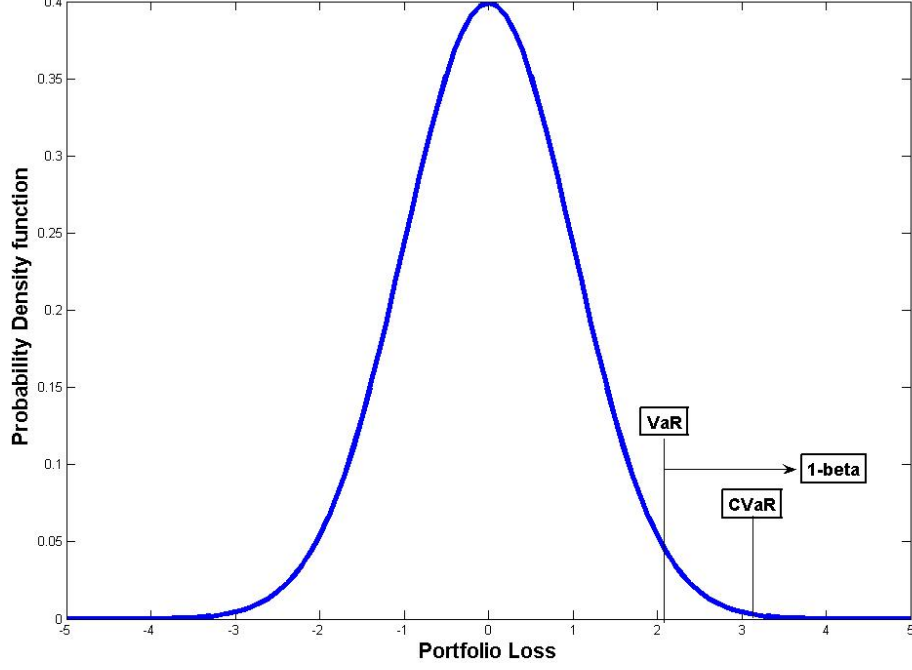


Figure 1.3: Portfolio loss distribution: VaR & CVaR

that is $\int_{f(x,\xi) \geq \alpha_\beta(x)} p(\xi) d\xi = 1 - \beta$. Then we have

$$\begin{aligned}
 & E[f(x, \xi) | f(x, \xi) \geq \alpha_\beta(x)] \\
 &= \int_{f(x,\xi) \geq \alpha_\beta(x)} f(x, \xi) p(\xi) d\xi \\
 &= \int_{f(x,\xi) \geq \alpha_\beta(x)} f(x, \xi) \frac{p(\xi)}{p(f(x, \xi) \geq \alpha_\beta(x))} d\xi \\
 &= \phi_\beta(x).
 \end{aligned} \tag{1.5}$$

β -CVaR is also called expected shortfall. This expectation must be greater than $\alpha_\beta(x)$, i.e. β -CVaR $\geq \beta$ -VaR. For an alternative way of defining CVaR, see [104].

VaR and CVaR (especially CVaR) have been widely analysed and applied in portfolio selection problems. The theory of probabilistic functions and percentiles was introduced in [122, 123], including the sensitivities of probabilistic functions, sensitivities of VaR, both with respect to decision variables, and the application of CVaR to portfolio optimisation modelling. The optimisation program with CVaR constraints was constructed in [77]. The application of CVaR in credit risk which is the exposition to counterparty default can be found in [4] based on CreditMetrics methodology. The problem with CVaR constraints was translated to L-shape and solved efficiently in [79]. A decomposition framework handling CVaR objectives and constraints in two-stage stochastic models was discussed in [41].

Although VaR is widely used as a risk measure nowadays, there are still some criticisms about its undesirable properties. The subadditivity which should be true for a coherent risk measure is not satisfied by VaR, which means the VaR of a portfolio with two instruments may be greater than the sum of the individual VaRs of the two instruments, i.e. $\alpha_\beta(x + \xi) \geq \alpha_\beta(x) + \alpha_\beta(\xi)$, see [6]. As we know that there is a covariance between two assets, this usually reduces the risk for portfolios consisting of the combination of two assets (other than two assets individually). Therefore, VaR is not coherent. In addition, it is difficult to optimise VaR in the discrete case. CVaR, as a coherent risk measure, has better properties than VaR:

1. Translation invariance: if we denote $\phi_\beta(x) = CVaR(f(x, \xi))$, where the linear function $f(x, \xi)$ is the amount of loss, then,

$$\phi_\beta(x + c) = \phi_\beta(x) + c.$$

It means that a set loss amount of c will lead to an increase of CVaR by c .

2. Subadditivity: $\phi_\beta(x_1 + x_2) \leq \phi_\beta(x_1) + \phi_\beta(x_2)$.
3. Positive homogeneity: for all $\lambda \geq 0$, $\phi_\beta(\lambda x) = \lambda \phi_\beta(x)$.
4. Monotonicity: for all x_1 and x_2 with $x_1 \leq x_2$, we have $\phi_\beta(x_2) \leq \phi_\beta(x_1)$.

In addition, for the same threshold value, since $CVaR \geq VaR$, if $CVaR$ is below this value, VaR is restricted to be less than this value as well. $CVaR$ can be considered as a replacement for VaR in this sense. Besides these properties, CVaR can also be linearised and solved by an LP solver [109], which is a great advantage in practice.

It is difficult to handle CVaR because of the VaR function $\alpha_\beta(x)$ involved in the definition, unless we have an analytical representation for VaR. However, β -CVaR(x, ξ) is proved in [109] to be equivalent in terms of the mean risk model to:

$$F_\beta(x, \alpha) = \alpha + (1 - \beta)^{-1} \int_{f(x, \xi) \geq \alpha} (f(x, \xi) - \alpha) p(\xi) d\xi,$$

which is convex and continuously differentiable with respect to α , i.e.

$$\phi_\beta(x) = F_\beta(x, \alpha_\beta(x)) = \min_{\alpha \in \mathbb{R}} F_\beta(x, \alpha).$$

According to this, it can be easily derived that:

$$\min_{x \in X} \phi_\beta(x) = \min_{(x, \alpha) \in X \times \mathbb{R}} F_\beta(x, \alpha),$$

where, moreover, a pair (x^*, α^*) achieves the right-hand side minimum if and only if x^* achieves the first minimum and α^* is the corresponding VaR. Hence, the mean-risk

model with CVaR can be constructed as:

$$\min_{(x,\alpha) \in X \times \mathbb{R}} -E[r'x] + F_\beta(x, \alpha).$$

And in the case that the random variables have discrete probability distributions, the integration in function $F_\beta(x, \alpha)$ becomes a summation. As a result, this optimisation problem becomes linear if the performance function $f(x, \xi)$ is linear.

The mean-risk model with CVaR as the risk measure is proven to be consistent with second-order stochastic dominance as shown in Section 3.2.3. While CVaR is concerned with one threshold value VaR, second-order stochastic dominance actually can be interpreted as a series of CVaR constraints for various threshold values.

1.2.4 Asset-Liability Management

Asset-Liability Management (ALM) is one main category of portfolio selection problems. Institutions involving large amounts of liquidity, like banks or insurance companies, seek out efficient portfolio strategies for the use of their assets and liabilities, under the consideration of the inherent uncertainty of portfolio returns, cash flows and consequent costs. The aim is to maximise the profit while satisfying liability in the mean time. The liabilities may take different forms: pensions paid to the members of the scheme in a pension fund, savers' deposits paid back in a bank, or benefits paid to insurees in the insurance company. The situation in which liability is not satisfied is called underfunding. To avoid underfunding while the return of assets is unpredictable is actually a crucial question in optimally allocating the assets. While the general portfolio selection problems considers the risk of market value decrease of assets, ALM is additionally concerned with the risk of underfunding. This work will specialise in the market risk involved in ALM.

Stochastic dominance, which is used to justify the efficiency of risk measures, can provide a partial rank of portfolios consistent with general utility functions and can be used to filter out the portfolios overperforming certain benchmarks that are set as a threshold. In Chapter 3 we will show how this can be done for ALM problems to achieve optimal portfolio strategy considering liabilities. Meanwhile, a specific issue in ALM is the risk of underfunding, which is the situation where liabilities can not be satisfied. The chance of underfunding should be kept under control, for which we introduce a variation of second-order stochastic dominance called relaxed interval second-order stochastic dominance, and show that this can be handled using linear constraints.

1.3 Operational Risk

The other category of financial risk we deal with in this project is operational risk, which is a fairly new topic and has limited sophisticated research outcomes. Operational risk management has been attracting attention and interest since the 1990s, when the fast

development and increasing complexity of the financial market led to more and more loss which was not due to market risk or credit risk, but instead was due to operational risk. Operational risk covers a wide range of events and can result from various sources, where managerial issues normally play significant parts. It can result from human or machines; can be hidden in calculations or documentations; can be external or internal. Before managing operational risk efficiently, it is important to identify what operational risk is.

1.3.1 Definition of Operational Risk

Three ways of defining operational risk are listed in [69]. In the broadest sense, operational risk is *any financial risk other than market risk*. This definition is fuzzy and too broad and makes it difficult to measure and control the risk. The narrowest approach defines operational risk as *risk arising from operations*, including back office problems, (technology) failures in transaction processing and in systems, and technology breakdown. However, some aspects other than operations should also be taken into account, e.g. internal fraud, improper sale practices or model risk. An intermediate and more acceptable definition is given by the Basel Committee on Banking Supervision in Basel II [87, 88]:

“the risk of loss resulting from inadequate or failed procedures, people, systems or from external events”.

Operational risk under this definition can be classified by sources and causes as in Table 1.1 [87, 88].

New stringent norms and regulations are in operation to help with operational risk management. But on the other hand, those new regulations make operations more complicated. The Sarbannes-Oxley Act of 2002, the most significant piece of securities legislation since the 1930s in the US, is one of them. By improving the system of financial reporting and reinforcing the responsibility and governance of corporate management which are critical to investor confidence, the Act aimed to better align the incentives of managers, auditors and other professionals with those of investors, especially in the wake of significant fraud at that time, e.g. the Enron scandal in 2001. It came into force in 2002 and applied to all corporations in the US.

1.3.2 Measuring and Managing Operational Risk

The logic underlying risk management in a Basel II context is always the same for each category of banking risks. The first pillar of Basel II proposes two methods to deduce the minimum capital amount: the standardised method (a simple way of calculation provided by Basel II that can be applied to most banks) and the advanced method (comprehensive ways involving more analysis and numerical tools developed for specific banks). When the bank invests in qualified staff, software, and develops an

Internal Risks		
People	Processes	Systems
Employee collusion/fraud	Accounting error	Data quality
Employee error	Capacity risk	Programming errors
Employee misdeed	Contract risk	Security breach
Employers liability	Misselling/suitability	Strategic risks
Employment law	Product complexity	(platform/suppliers)
Health and safety	Project risk	System capacity
Industrial action	Reporting error	System compatibility
Lack of knowledge skills	Settlement/payment error	System delivery
Loss of key personnel	Transaction error	System failure
	Valuation error	System suitability

External Risks	
External	Physical
Legal	Fire
Money laundering	Natural disaster
Outsourcing	Physical security
Political	Terrorist
Regulatory	Theft
Supplier risk	
Tax	

Table 1.1: Source: British Bankers' Association survey.

advanced model, the bank is able to “economise” some capital assuming that it contributes actively to risk management. The common factor of these models is that they rely on quantitative variables whose behaviour is described by theoretical probability distributions. These parametric methods have the advantage of being defined by just a few parameters (e.g. centrality and dispersion indicators for the normal probability distribution function). Just to name a few of them, the mean-variance framework (i.e. based on the Markowitz model), the Sharpe ratio (i.e. CAPM model), the different VaR (except for the VaR based on historical simulations), are all examples of such methods. When the bank is not up to developing its own advanced model, then the capital that is to be set aside is calculated using a standardised approach.

Several approaches have been developed to calculate the minimum amount of capital required for managing operational risk, corresponding to the advanced measurement approach. For instance, Panjer [100] exploited probability theory and statistical tools that could help build up operational risk management models. Generally, these approaches can be categorised as belonging to one of two methodologies: top-down and bottom-up [24, 90, 130]. The top-down methodology measures the operational risk as a percentage of a certain amount of profit or other quantity without analysing the causes of risk. In contrast, the bottom-up methodology researches on the probabilities and severities of events which could lead to loss in order to measure the risk. In this

methodology, there are statistical methods (estimating probability distribution of risk events), causal approaches (based on cause-effect analysis) and Bayesian approaches (via Bayes theorem). For example, [82] proposes a statistical model combined with cause-effect analysis named the OpRisk Tree model. In this method, people firstly identify the risk events and classify them; they then draw a flow diagram similar to a decision tree to demonstrate the cause and effects between risk events and control behaviour; thirdly, they assign probability distributions in this diagram; a simulation of the frequency and severity of each risk event is then run, and then an analysis of the correlations of these events can be conducted. There is also operational risk analysis in specific areas: complexity introduced by derivative products in [64] and operational risk in foreign exchange market analysed in [74]. Both methodologies can deduce the amount of money to be set aside to cover the possible loss in the future due to operational risk.

The book of Cruz [25] presents a general framework for operational risk modelling, managing and hedging. The modelling and measuring part covers three aspects: data modelling (database construction, key indicators for operational risk), stochastic modelling (using statistical theory to model the severity and frequency of operational risk events as two discrete stochastic processes) and casual modelling (analysing cause and effects of operational risk events). The author uses operational Value-at-Risk to measure the risk modelled in a stochastic manner. Operational risk management is proposed in an approach composed of separate reports to regulators and managers, real-time control and cost control. To hedge operational risk, it is suggested using risk mitigation by derivatives, insurance and capital allocation.

However, measuring operational risk is rather subjective due to its qualitative nature, being related as it is to managerial issues. The benefit from these methodologies, which heavily rely upon statistical theory and quantitative variables to measure and manage operational risk is really limited. Instead of using one single tool to digest all aspects and to generate rough management strategy, in the following section we propose a stochastic programming model to deal with the operational risk of human origin. The method is illustrated by a Workforce Planning Problem.

1.3.3 Workforce Planning Problem

It is common in financial sectors that both professionals and non-professionals are employed to deliver service. Much of the work involved in the financial industry demands people with high education, experimental knowledge and certain skills, while other parts can be dealt with by routine processes or under simple instructions. A notion arising from service science makes the distinction between explicit (information) and implicit (or tacit) knowledge [102]. Explicit knowledge can be recorded on documents easily and learnt by most people within short period of time. Such knowledge is sufficient for employees dealing with routine work. Implicit knowledge, corresponding to professional knowledge and skills, requires much more time to understand and to be

put into practice. This kind of knowledge may not be requested on a daily basis, but it is critical to maintain a high quality of business. Especially when operations go into unusual situations for which explicit knowledge does not show the solution, only professionals with implicit knowledge are able to drive it back to the right track. From the point of view of capacities of employees, professionals are always preferable to non-professionals. On the other hand, however, the cost of hiring professionals is higher than the other group. It is a management decision to plan the workforce optimally considering the operational risk and the resulting benefit.

The Aggregate Planning Model (APM) adapted to services can be successfully applied to assess the level of expertise necessary to deal with operational risks in the back offices of banks. The Aggregate Planning Model was developed in the middle of last century and has been successfully applied in production planning problems, see [58, 78, 114], and manpower planning problems, see [1, 38, 75]. The most important feature of the Aggregate Planning Model is the aggregation, either of products or manpower or both, which are presented as inventory constraints. We demonstrate it using different extensions of the stochastic programming version of the Aggregate Planning Model. The novelty of approach presented in this research is to apply APM to *services* rather than to *goods*. Indeed, the “production” of services is intangible (see [101] for the characteristics of service). In other words, services do not produce goods and therefore cannot contain inventory of goods. In the APM approach presented in this research, we consider the *inventory of human expertise* available in the back office.

Therefore, instead of taking all aspects of operational risk together and dealing with it ambiguously, we prefer to analyse one source of operational risk, i.e. employees, which results in a large portion of the loss from operational risk, using the Aggregate Planning Model framework. Starting from a classical multistage linear programming version of the Aggregate Planning Model, in Chapter 4 we develop several extensions that enable us to capture the true nature of operations risks.

1.4 Stochastic Programming

The operational research methodology, as it has developed over many decades, has been serving the world in many aspects. Using the methodology, people observe and formulate real problems carefully, then model those problems in scientific way, by considering viewpoints as wide and reasonable as possible. Then a solution is obtained to optimality. Those features of operational research characterizes its applications in a variety of fields.

Probability Theory and statistics help people better understand and recognise the rules by which systems operate. Generally, events will happen with probabilities. That different outcomes will occur under the same circumstances is explained by the fact that those outcomes all have positive probabilities. In many cases, it is more precise to predict what the future will more likely be rather than what the future will exactly be.

A combination of operational research with statistics results in stochastic programming. Stochastic programming is mathematical programming with random parameters. In a real context of enterprise risk management, when future events need to be considered in business activity planning, uncertainty of parameters plays the key role. Initiated in the late fifties by Dantzig and Madansky, stochastic programming provides a paradigm to include uncertainty into optimisation-based decision models [13, 71]. In particular, a multistage stochastic program with recourse is a multi-period mathematical program where parameters are assumed to be uncertain along the time path. The term recourse means that the decision variables adapt to the different outcomes of random parameters at each time period. The stochastic programming model allows one to handle several scenarios simultaneously while scenarios present different outcomes of random variables. It provides an adaptive policy that is close in spirit to the way decision-makers have to deal with uncertain futures in real life.

Uncertainty, which is everywhere in the financial market, is the common characteristic of market risk management and operational risk management. It follows that stochastic programming is suited well to both risk management problems. In managing market risk in ALM problems, the uncertainty of asset return is unavoidable. A mean-risk framework is adopted with reasonable risk control to generate optimal portfolio strategy whose risk is tolerable. In managing operational risk resulting from human factors, the uncertainty of service demand and efficiency of employees cannot be ignored. A stochastic Aggregate Planning Model is built up to allocate the workforce optimally so that a service of high quality is guaranteed with lowest cost.

However, there are also differences in dealing with market risk and operational risk. As a result, different techniques are needed in constructing stochastic programming models.

Stochastic programming makes modelling possible in the case the parameters needed are random, i.e. the value could be from sets, continuous or discrete. We still need to know what set it is and the corresponding behavior (probability distribution over this set). A big assumption of general stochastic programming is that the probability distributions of random parameters are known. In most cases, we can use historical data or do simulation with assumptions upon the statistical parameters. We will use historical market data in market risk management. However, this is rarely applicable in operational risk modelling, where historical data is limited and statistical parameters are complicated to evaluate. Therefore, we propose the use of Bayesian network theory to construct the probability distribution of random parameters in the programming. This will be illustrated in Section 4.4.

Most stochastic programming used today is in the context of exogenous problems, where uncertainty is independent of decision variables and can be predicted as a set of values in advance. An individual investor's trading performance can rarely influence the whole market. We also assume that institutional trading action would not lead significant changes to the market. This means that our portfolio strategy would not

have any subsequent impact on the behavior of the market, i.e. the parameters in the model are independent on our decisions. For other problems, endogenous random factors exist, where the decision taken at current stage will influence the behaviour of those uncertainties. This will be an important issue to be taken into account in modelling operational risk resulting from human factors. Section 4.5 will show how to model this fact in a stochastic Aggregate Planning Model.

The other issue in stochastic programming besides modelling comes from the solution techniques. The main challenge in solving stochastic programming is the size of the model. It can easily grow with the increase of time horizons and the set of random parameters values. There are algorithms, like Benders decomposition method and progressive hedging algorithm [108], which take the advantage of the special structure of typical stochastic programming models. A certain number of stochastic programming models, however, do not contain such structure. As a result, a direct decomposition method is not as efficient as for those typical stochastic programming models. An object-oriented parallel solver (OOPS), based on interior point method, is employed to solve the ALM models constructed here. By exploiting the special structure present in the models, this solver works in an efficient manner saving both computation time and memory requirement.

This research studies two modelling techniques based on stochastic programming that help seek optimal strategies in financial risk management. The first technique is concerned with market risk management for an Asset and Liability Management problem. A stochastic dominance concept is applied to control the market risk and the risk of underfunding. A small numerical example and an out-of-sample backtest demonstrate advantages of this new model. By creating a link between variables associated with different scenarios of the same time stage, the risk control constraints always destroy the usual tree-structure of the constraint matrix in the stochastic program and prevent the application of standard stochastic programming approaches such as (nested) Benders decomposition or progressive hedging. A structure-exploiting solver (OOPS) is applied to this problem. Computational results on medium scale problems with sizes reaching about one million variables demonstrate the efficiency of the specialised solution technique.

The second model deals with operational risk from human origin by optimally allocating workforce from a managerial perspective. Unlike market risk that can be handled in a financial manner, the treatment of operational risks calls for a “managerial approach”. Consequently, we propose a new way of dealing with operational risk, which relies on the well known Aggregate Planning Model. Our contribution corresponds to several improvements applied to stochastic programming modelling. First, the incorporation of randomness associated with demand volume and with the capacity of work provided by employees transforms the general APM into a multistage stochastic program. Then, as advocated by Basel II, we calculate the probability distribution based on a Bayesian Network to circumvent the difficulty of obtaining data

which characterises uncertainty in operations. Third, we go a step further by relaxing the traditional assumption in stochastic programming that imposes a strict independence between the decision variables and the random elements. Comparative results show that in general these improved stochastic programming models tend to allocate more human expertise in order to hedge operational risks. The dual solutions of the stochastic programs are exploited to detect periods and nodes that are at risk in terms of expertise availability.

1.5 Outline of the Thesis

In Chapter 2, we introduce and formalise stochastic programming methodology. Our focus will be upon the recourse problem, from 2-stage problem to multiple-stage problem. The block-angular structure of the problem will also be described. Corresponding to this special structure, we will present and compare two algorithms dealing with stochastic programming models, i.e. a dual decomposition method and an interior-point method based parallel solver. At the end of this chapter, endogenous problems will be discussed briefly, followed with a list of applications of stochastic programming.

Asset-Liability Management is detailed in Chapter 3, where we use stochastic dominance to control the market risk. The modelling methodology is explained and theoretical issues are also discussed. Furthermore, we introduce a variation of second-order stochastic dominance and develop chance constraints from this variation. A small numerical example and an out-of-sample backtest will show how this model works. Due to the challenge arising from solving stochastic programming models, OOPS, an interior-point solver, will be used to solve the ALM models built up and its performance will be illustrated in the implementation part.

In Chapter 4, operational risk originating from human factors is analysed and modelled as a workforce planning problem according to Aggregate Planning Model framework. While ALM modelling follows a general stochastic programming prototype, this workforce planning model breaks through the two major assumptions made in general stochastic programming regarding to the uncertain factors. To model the risk precisely, we take three steps starting from the stochastic Aggregate Planning Model with random service demand, i.e. firstly considering the randomness of operation efficiency, secondly revising the probability distribution of operation efficiency using a Bayesian network, and thirdly adapting the influence of decision variables on the uncertainty of demand upon the model. The implementation is illustrated for each stage of development, followed by a perspective from the point view of shadow prices.

Chapter 5 concludes the thesis and proposes prospective research directions.

Chapter 2

Some Modelling and Solution Techniques of Stochastic Programming

Stochastic programming has been identified as a useful tool dealing with uncertainties in optimisation problems for more than half a century. In this chapter, we will first introduce the general concepts of stochastic programming following the presentation of Kall and Wallace [71] and then focus on recourse problems. We then compare two solution methods of Benders decomposition and an interior-point method based algorithm, followed with a brief discussion of endogenous and exogenous problems. A list of applications of stochastic programming is presented at the end of this chapter.

2.1 Induction of Stochastic Programming

Optimisation is often used to model real problems under the strong assumption that all the data is known. However, for some real-life problems it is either too expensive or even impossible to guarantee this certainty, in which case stochastic programming is explored.

Many practical problems, assuming known parameters, can be modeled as a mathematical programming problem:

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \dots, m \\ & x \in X \subset \mathbb{R}^n, \end{aligned} \tag{2.1}$$

where the set X and functions $f_0, g_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, m$, are generated during the modelling process.

In real-life applications many coefficients, i.e. the parameters in f_0 and $g_i, i = 1, \dots, m$, are unpredictable, in which case, models of fixed coefficients may be impre-

cise or even not the problem that we want to deal with. Such coefficients should be considered uncertain or random. Stochastic programming takes this uncertainty into account and aims to find the optimal solution to the problems involving uncertainties. Allowing the uncertainty of parameters in (2.1), the stochastic program can be presented as:

$$\begin{aligned} \text{“min”} \quad & f_0(x, \bar{\xi}) \\ \text{s.t.} \quad & g_i(x, \bar{\xi}) \leq 0, \quad i = 1, \dots, m \\ & x \in X \subset \mathbb{R}^n, \end{aligned} \tag{2.2}$$

where $\bar{\xi}$ is a random variable varying over a set $\Omega \subset \mathbb{R}^k$ with probability measure \mathcal{P} , or say in the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where \mathcal{F} is an algebra over set Ω and \mathcal{P} is the probability measure of \mathcal{F} . As a result, we can see that the values of functions $f_0(x, \cdot)$ and $g_i(x, \cdot)$, $i = 1, \dots, m$, are random as well.

The precise characterisation of the feasible set, as well as the question of how to perform the minimisation of the objective function in (2.2), has yet to be clearly defined, that's why there are quotation marks on the minimisation. With different realisations of the random factor $\bar{\xi}$, the objective function and constraint functions have different values. Before we solve this problem, a further step to revise modelling is necessary and this leads to an equivalent deterministic programme. There are several ways to do this revision, e.g. consider the worst case, or consider the best case, or take expectation over all random values. In this work, we will follow the way of stochastic programming with recourse as illustrated in the next section.

2.2 Recourse Problems

The presentation in this section follows closely Kall and Wallace [71].

We introduce the deterministic equivalent to (2.2) in the manner of inclusion of recourse. Denote

$$g_i^+(x, \xi) = \begin{cases} 0, & \text{if } g_i(x, \xi) \leq 0 \\ g_i(x, \xi), & \text{otherwise} \end{cases}$$

Then i th constraint of (2.2) is violated if and only if $g_i(x, \xi) > 0$ for a given decision x and realisation ξ of $\bar{\xi}$. Hence, to compensate for this violation after observing the value of ξ , we could add a recourse variable $y_i(\xi)$ to each corresponding constraint such that $g_i(x, \xi) - y_i(\xi) \leq 0$. For instance in the production planning problem, when there is a shortage of products so that demand cannot be fully satisfied, purchase can be operated from other suppliers to make it up. This compensation is assumed to be accompanied by an extra cost or penalty which is q_i per unit. These additional costs

can be formulated as the recourse functions and amount to:

$$Q(x, \xi) = \min_y \left\{ \sum_{i=1}^m q_i y_i(\xi) \mid y_i(\xi) \geq g_i^+(x, \xi), \ i = 1, \dots, m \right\}, \quad (2.3)$$

yielding a total cost of

$$f(x, \xi) = f_0(x, \xi) + Q(x, \xi).$$

A more general linear recourse program with a recourse vector $y(\xi) \in Y \subset \mathbb{R}^k$, where Y is some given polyhedral set, and an arbitrary fixed $m \times k$ matrix W , which is referred to as the recourse matrix, can be formulated as follows:

$$Q(x, \xi) = \min_y \{ q'y \mid Wy \geq g^+(x, \xi), \ y \in Y \}, \quad (2.4)$$

where $q \in \mathbb{R}^k$ is a corresponding unit cost vector and $g^+(x, \xi) = (g_1^+(x, \xi), \dots, g_m^+(x, \xi))$.

By constructing a reasonable recourse function using any of (2.3), (2.4), to achieve the minimal expected value of the total costs, model (2.2) can be revised as a deterministic equivalent of a stochastic program with recourse:

$$\min_{x \in X} E_{\bar{\xi} \in \Omega} f(x, \bar{\xi}) = \min_{x \in X} E_{\bar{\xi} \in \Omega} \{ f_0(x, \bar{\xi}) + Q(x, \bar{\xi}) \}. \quad (2.5)$$

This equivalent program is considered as a two-stage program, i.e. recourse is determined at the second stage after decision x made at the first stage and the value of random variable ξ known.

After generating the deterministic equivalent, one may be concerned with the mathematical properties of such programs, e.g. convexity, so as to determine if it is manageable with general algorithmic and computational capabilities. The following proposition traced back to Kall and Wallace [71] addresses this issue.

Proposition 1. *If $f_0(\cdot, \xi)$ and $Q(\cdot, \xi)$ are convex in X , $\forall \xi \in \Omega$, and if X is a convex set, then (2.5) is a convex program.*

In the next section, we will specifically focus on the linear stochastic programs when the recourse (2.4) is linear.

2.2.1 Linear Recourse Programming

Linear recourse is relevant to most practical problems and has favourable properties in terms of solution methods.

Consider a linear stochastic program,

$$\begin{aligned} \text{“min”} \quad & c'x \\ \text{s.t.} \quad & Ax = b, \\ & L(\bar{\xi})x = h(\bar{\xi}), \\ & x \geq 0, \end{aligned} \quad (2.6)$$

where both $A \in \mathbb{R}^{m_0 \times n}$ and $b \in \mathbb{R}^{m_0}$ are deterministic, the matrix $L(\cdot) \in \mathbb{R}^{m_1 \times n}$ and vector $h(\cdot) \in \mathbb{R}^{m_1}$ in contrast depend on the random variable $\bar{\xi}$. The set X in formulation (2.2) is defined as $X = \{x \mid Ax = b, x \geq 0\}$ in this case.

Following the approach of (2.5) and (2.4) in constructing a deterministic equivalent to the stochastic program (2.6), we have the following stochastic linear program with linear recourse:

$$\begin{aligned} \min_x \quad & c'x + E_{\bar{\xi}}\{Q(x, \bar{\xi})\} \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0, \end{aligned}$$

where

$$Q(x, \xi) = \min_y \{q'y \mid W(\xi)y = h(\xi) - L(\xi)x, \quad y \geq 0\}, \quad (2.7)$$

and the recourse matrix $W(\xi) \in \mathbb{R}^{m_1 \times n}$ depends on the random variable $\bar{\xi}$.

The second-stage decision variable y in the stochastic program is often used to model the compensation of deficiencies arising in practical applications. It is expected that, given any first-stage decision $x \in X$ and any realisation of random variable $\bar{\xi}$, such compensation exists, i.e. the second-stage program

$$\begin{aligned} Q(x, \xi) = \min_y \quad & q'y \\ \text{s.t.} \quad & W(\xi)y = h(\xi) - L(\xi)x, \\ & y \geq 0, \end{aligned} \quad (2.8)$$

is feasible $\forall \xi \in \Omega$. To specify the conditions of feasibility, we assume a discrete distributed random variable $\bar{\xi}$ with realisations ξ^j , $j = 1, \dots, r$. Then, the first-stage feasibility set K can be generated as

$$K = \{x \mid L(\xi^j)x + W(\xi^j)y^j = h(\xi^j), \quad y^j \geq 0, \quad j = 1, \dots, r\}.$$

For any $x \in X \cap K$, and realisation ξ , the recourse problem is always feasible. Hence, if $X \cap K \neq \emptyset$, the stochastic program is feasible. We now introduce the concept of complete fixed recourse. A recourse matrix W is fixed if it does not depend on the random variable $\bar{\xi}$. Furthermore, a fixed $m_1 \times n$ recourse matrix W is defined to be complete if

$$\{z \mid z = Wy, \quad y \geq 0\} = \mathbb{R}^{m_1},$$

implying that $K = \mathbb{R}^n$, i.e. for any first-stage decision x and any realisation ξ of $\bar{\xi}$, the second-stage program (2.8) is always feasible.

In fact, relatively complete recourse is sufficient for feasibility of the second-stage

program, which is defined as

$$h(\xi) - L(\xi)x \in \{z \mid z = Wy, y \geq 0\}, \quad \forall \xi \text{ and } \forall x \geq 0 \text{ satisfying } Ax = b.$$

It is obvious that complete recourse implies relatively complete recourse.

Furthermore, if the random variables have a finite joint discrete probability distribution

$$P(\bar{\xi} = \xi^k) = p_k, \quad \sum_{k=1}^r p_k = 1, \quad p_k \geq 0, \quad k = 1, \dots, r,$$

the expected value of the recourse function can be explicitly stated. We can rewrite the stochastic program (2.7) as

$$\begin{aligned} \min_x \quad & c'x + \sum_{k=1}^r p_k q' y^k \\ \text{s.t.} \quad & Ax = b \\ & L(\xi^k)x + Wy^k = h(\xi^k), \quad k = 1, \dots, r, \\ & x \geq 0, \\ & y^k \geq 0, \quad k = 1, \dots, r. \end{aligned} \tag{2.9}$$

The two-stage stochastic linear program has the special structure named *dual decomposition structure* as illustrated in Figure 2.1. The block structure is well-suited to the *dual decomposition* method discussed in Section 2.4.2.

2.2.2 Multiple-Stage Recourse Problems

It is natural to extend the two-stage recourse program discussed in the previous section to a multi-stage context by considering decisions in more stages other than x at the first stage and y at the second stage.

Assume there are $T + 1$ stages with $T + 1$ sequential decisions x_τ to be taken at stages $\tau = 0, 1, \dots, T$. Stages here do not necessarily have to correspond to time periods. Then, at stage τ ($\tau \geq 1$), knowing the previous decisions and values of random variables, including realisations ξ_1, \dots, ξ_τ of $\bar{\xi}_1, \dots, \bar{\xi}_\tau$ and decisions $x_0, \dots, x_{\tau-1}$, we have to decide on x_τ satisfying the constraints:

$$g_\tau(x_0, \dots, x_\tau, \xi_1, \dots, \xi_\tau) \leq 0,$$

where g_τ are vector-valued functions.

Assuming a cost function $q_\tau(x_\tau)$ at stage $1 \leq \tau < T$ which are known parameters at the beginning, the recourse function can be presented recursively as

$$\begin{aligned} Q_\tau(x_0, x_1, \dots, x_{\tau-1}, \xi_1, \dots, \xi_\tau) = \min_{x_\tau} \quad & E_{\bar{\xi}_{\tau+1}} \{q_\tau(x_\tau) + Q_{\tau+1}(x_0, \dots, x_\tau, \xi_1, \dots, \xi_\tau, \bar{\xi}_{\tau+1})\} \\ \text{s.t.} \quad & g_\tau(x_0, \dots, x_\tau, \xi_1, \dots, \xi_\tau) \leq 0. \end{aligned}$$

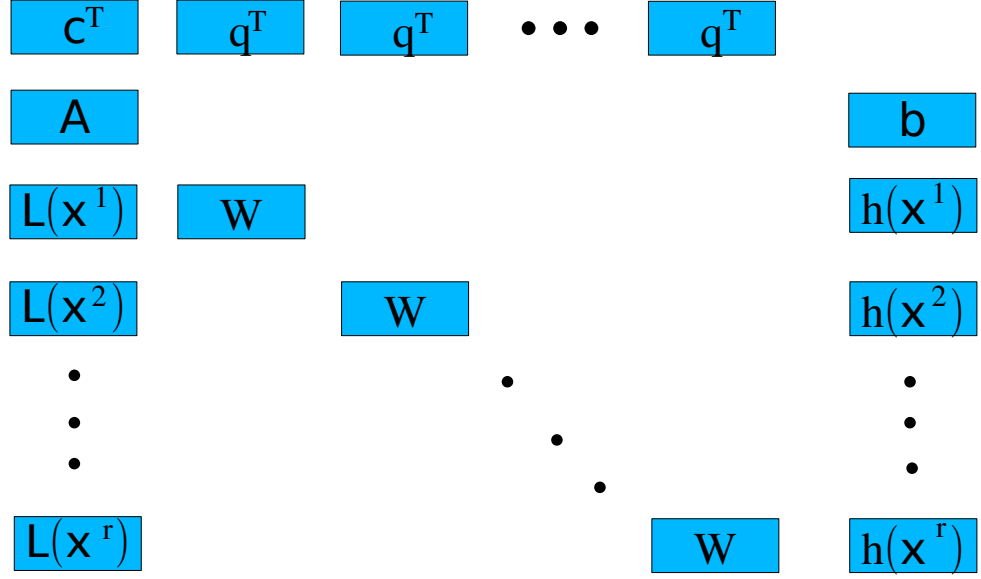


Figure 2.1: Block structure of a two-stage recourse problem.

And at stage T ,

$$Q_T(x_0, x_1, \dots, x_{T-1}, \xi_1, \dots, \xi_T) = \min_{x_T} \{q_T(x_T) \mid g_T(x_0, \dots, x_T, \xi_1, \dots, \xi_T) \leq 0\}.$$

It can be easily seen that the optimal recourse decision \hat{x}_τ depends on $x_0, \dots, x_{\tau-1}$ and ξ_1, \dots, ξ_τ , i.e.

$$\hat{x}_\tau = \hat{x}_\tau(x_0, x_1, \dots, x_{\tau-1}, \xi_1, \dots, \xi_\tau), \quad \tau \geq 1.$$

The deterministic equivalent for this multi-stage stochastic program with recourse, which involves dynamic decision process, can be written as

$$\min_{x_0 \in X} E_{\bar{\xi}_1} \{f_0(x_0) + Q_1(x_0, \bar{\xi}_1)\},$$

which is similar to (2.5) and a straightforward extension for multi-stage problems. Proposition 1 is also true for multi-stage case.

The evolution of uncertainties in multi-stage stochastic programming in the discrete case can be presented as an event tree, which describes the unfolding of values of the random parameters through the time horizon considered. Take a 3-stage problem for example as shown in Figure 2.2. Each path from the root to a leaf node is a scenario presenting a possible outcome of the uncertainties, which is assigned with a probability.

Each node of the event tree is defined with a set of constraints, a set of variables and its ancestor node. The associated decisions at a non-leaf node are the same to all the scenarios sharing this node, at the corresponding stage of that node.

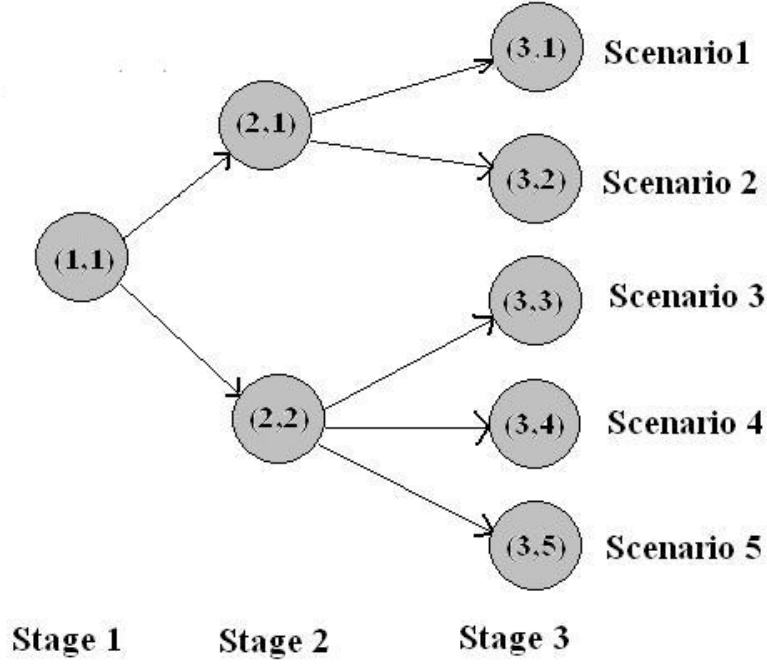


Figure 2.2: A simple event tree.

Similar to the 2-stage problem, a multi-stage stochastic linear program with linear recourse has the following formulation:

$$\begin{aligned}
 \min_x \quad & E_{\bar{\xi}_1} \{c'x_0 + Q_1(x_0, \bar{\xi}_1)\} \\
 \text{s.t.} \quad & Ax_0 = b \\
 & x_0, \dots, x_T \geq 0,
 \end{aligned} \tag{2.10}$$

where

$$\begin{aligned}
 Q_\tau(x_0, \dots, x_{\tau-1}, \xi_1, \dots, \xi_\tau) &= \min_{x_\tau} E_{\bar{\xi}_{\tau+1}} \{q'_\tau x_\tau + Q_{\tau+1}(x_0, \dots, x_\tau, \xi_1, \dots, \xi_\tau, \bar{\xi}_{\tau+1})\} \\
 \text{s.t.} \quad & W(\xi_\tau)x_\tau = h(\xi_\tau) - L(\xi_\tau)x_{\tau-1}.
 \end{aligned}$$

Furthermore, if the random variables have a finite joint discrete probability distribution, i.e. r_τ nodes at stage τ , the above stochastic program can be rewritten

as

$$\begin{aligned}
\min_x \quad & c'x_0 + \sum_{\tau=1}^T \sum_{k_\tau=1}^{r_\tau} p^{k_\tau} q'_\tau x_\tau^{k_\tau} \\
\text{s.t.} \quad & Ax_0 = b \\
& L(\xi^{k_\tau})x_{\tau-1}^{k_{\tau-1}} + W(\xi^{k_\tau})x_\tau^{k_\tau} = h(\xi^{k_\tau}), \quad \tau = 1, \dots, T, \quad k_\tau = 1, \dots, r_\tau, \\
& x_0, \dots, x_T \geq 0.
\end{aligned} \tag{2.11}$$

By associating an event tree to the program and listing the constraints by depth-first search ordering of the nodes, the constraint matrix corresponding to the event tree of Figure 2.2 has a nested dual block-angular structure as presented in Figure 2.3, similar to Figure 2.1.

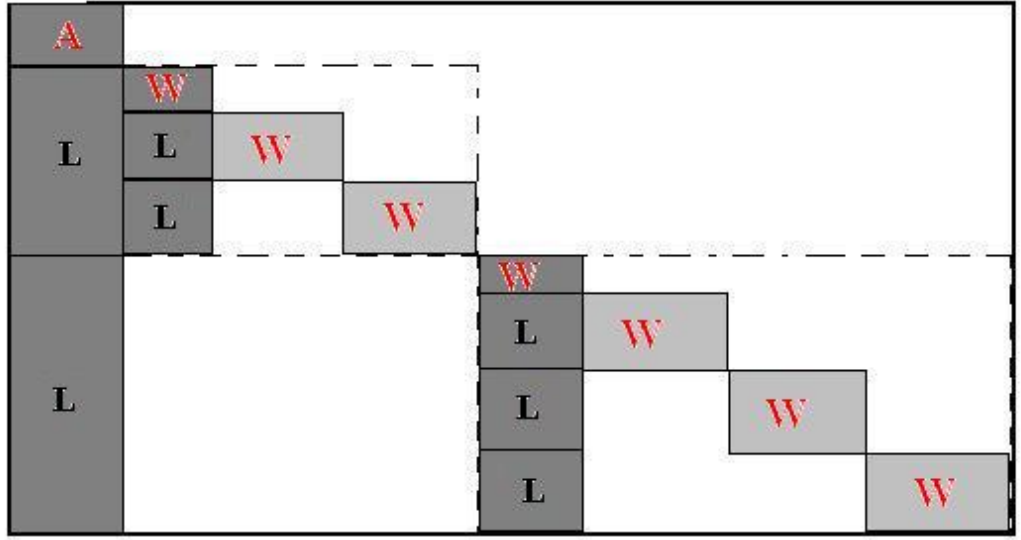


Figure 2.3: Block structure of a multi-stage linear recourse problem.

Both the ALM problem and workforce planning problem considered in this project are modelled as multi-stage programs with linear recourse. While the ALM modelling is linear, the workforce planning, however, is a mixed-integer program where the objective function and constraint functions are linear but the decision variables need to be integers. The solution techniques for stochastic linear programming will be discussed in a later section. How to solve the stochastic integer programming is out of the scope of this work. We will resort to reliable integer solvers for the workforce planning models introduced in Chapter 4.

2.3 Scenario Generation

In the previous section, we mentioned the event tree which describes the unfolding of values of random parameters through the modelling horizon. When modelling a stochastic problem, it is a very important issue to estimate the random parameters. Scenario can be applied to present realizations of random factors under the assumption that the random variables have discrete probability distributions. The corresponding probability distribution could be analyzed and concluded from historical data. Then a scenario tree can be generated from the probability distribution [16], i.e. by sampling values from the probability distribution of ξ . If there are correlated random variables, it would be necessary to specify the marginal distributions and the correlation matrix [15]. In case of unknown probability distribution, scenarios could be generated with required moments [117], e.g. mean, variance, skewness, etc. However, there is danger that the scenario tree grows exponentially and leads to difficulty in solving the model, when scenario reduction techniques [35] can be applied to reduce the size of the tree with the minimum loss of accuracy. Comparison of scenario generation methods can be found in [73].

2.4 Solving Stochastic Linear Programming

The main challenge in stochastic programming is the size of the deterministic equivalent, especially when there are more than two stages. The number of contingent variables and constraints grows linearly with the number of realisations of random factors and exponentially with the number of stages. The size of the problem creates a challenge to the solution approach. On the other hand, stochastic linear programming models present a special structure (as shown in Figures 2.1 and 2.3) that can be exploited to improve the efficiency of the solution approach. The general approach is to resort to a decomposition principle. Two solution techniques are discussed below, namely the Benders decomposition and an interior-point method based approach which consists of decomposing the linear algebra operations.

2.4.1 Duality Theory and Shadow Prices

Before introducing the decomposition methods, we review some basic statements from duality theory for linear programming. Given a linear program, which is also called the primal program, of the form

$$\begin{aligned} \min \quad & c'x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0, \end{aligned} \tag{2.12}$$

the so-called dual program can be formulated as

$$\begin{aligned} \max \quad & b'u \\ \text{s.t.} \quad & A'u \leq c. \end{aligned} \tag{2.13}$$

The feasible set of the primal program is denoted by $\mathcal{P} = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ and that of its dual by $\mathcal{D} = \{u \in \mathbb{R}^m : A'u \leq c\}$. We say

$$\begin{aligned} \inf_{x \in \mathcal{P}} c'x &= +\infty, \quad \text{if } \mathcal{P} = \emptyset, \\ \sup_{u \in \mathcal{D}} b'u &= -\infty, \quad \text{if } \mathcal{D} = \emptyset. \end{aligned}$$

It is proven that

$$\inf_{x \in \mathcal{P}} c'x \geq \sup_{u \in \mathcal{D}} b'u, \tag{2.14}$$

which is called weak duality. In other words, the dual objective provides a lower bound on the primal objective, and the primal provides an upper bound on the dual.

Furthermore, if either $\mathcal{P} \neq \emptyset$ or $\mathcal{D} \neq \emptyset$ then it follows that [71]:

$$\inf_{x \in \mathcal{P}} c'x = \sup_{u \in \mathcal{D}} b'u.$$

If one of these two problems is solvable then so is the other, and we have

$$\min_{x \in \mathcal{P}} c'x = \max_{u \in \mathcal{D}} b'u.$$

By the strong duality at the optimum, if both the primal and dual problems are feasible, the optimal value is then

$$z = c'\hat{x} = b'\hat{u}, \tag{2.15}$$

where $\hat{x} \in \mathbb{R}^n$ and $\hat{u} \in \mathbb{R}^m$ are the optimal solutions to the primal and dual problems, respectively.

From (2.15), we can see that if the dual solution does not change, an increase of b_j , $j = 1, \dots, m$ by 1 can lead to a \hat{u}_j unit increase in the optimal value. The dual solution \hat{u} represents the so-called shadow prices. The definition of a shadow price can be found in [65, 125]:

“The shadow price of the i th constraint is the amount by which the optimal z -value is improved (increased in a max problem and decreased in a min problem) if we increase b_j by 1 (from b_j to $b_j + 1$). ”

A “ \geq ” constraint will always have a nonpositive shadow price; a “ \leq ” constraint will always have a nonnegative shadow price; an equality constraint has a shadow price not restricted in sign.

A shadow price provides valuable information for the management. For instance in the production planning problem, there are recourse constraints, i.e. the right hand side b_j represents the amount of available j th material. While the dual solution remains optimal, increase of one more unit of b_j can bring \hat{u}_j more profit. This information is very helpful when an increase in j th material can be achieved only by simply purchasing more of the resource in the marketplace. This also implies the maximum price the manager would like to pay for an additional unit of a resource.

Furthermore, the shadow price is also related to sensitivity analysis of the linear program, whose main purpose is to identify the sensitive parameters (namely those that cannot be changed without changing the optimal solution). For those parameters, the value has to be estimated with special care, due to the risk of obtaining an erroneous optimal solution. $\hat{u}_j = 0$ implies that the optimal solution is not sensitive to at least small changes in b_j . On the other hand, for those \hat{u}_j 's with large absolute values, the optimal solution could be very sensitive to b_j and extra care is needed. The sensitivity analysis of nonlinear programs will be discussed in Chapter 4.

2.4.2 Benders Decomposition

The Benders decomposition algorithm [9] (also called the dual decomposition method) is one of the most popular algorithmic schemes to deal with two-stage linear stochastic programming problems. The use of the Benders decomposition method requires the dual decomposition structure of the problem where the coefficients of recourse variables (second-stage variables) appear as a block diagonal matrix as shown in Figure 2.1. Essentially, one considers a two-stage stochastic program and defines two independent problems, one for each stage. The problem decomposes at the second stage, so we have as many subproblems as there are second stage nodes. The decomposition algorithm consists of a dialogue between the problem at the first stage and subproblems at the second stage, in order to reach an optimum for the original problem. We have already seen that programs (2.9) and (2.11) have a dual decomposition structure; we now show how this special structure can be used in the Benders decomposition algorithm.

Restate the program (2.9) as

$$\begin{aligned} \min_x \quad & c'x + \sum_{k=1}^r f^k(x) \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0, \end{aligned} \tag{2.16}$$

where

$$f^k(x) = \min_{y^k} \{q' y^k \mid W y^k = h^k - L^k x, y^k \geq 0\}, \quad k = 1, \dots, r.$$

It can be obtained that the recourse function $f^k(x)$ is piecewise linear and convex. For any given value of x , recourse functions $f^k(x)$, $k = 1, \dots, r$, are independent. The

subproblems are r recourse programs $f^k(x)$, $k = 1, \dots, r$. Program (2.16) is equivalent to:

$$\begin{aligned} \min_{x, \theta^k} \quad & c'x + \sum_{k=1}^r \theta^k \\ \text{s.t.} \quad & Ax = b \\ & \theta^k - f^k(x) \geq 0, \quad k = 1, \dots, r, \\ & x \geq 0, \end{aligned} \tag{2.17}$$

with $f^k(x)$ defined above. (2.17) is called the master problem.

The decomposition method starts with a reasonable value \hat{x}_0 of x and $\hat{\theta}_0$ of θ , with which we evaluate the recourse function in the formulations of both primal and dual programs:

$$\begin{aligned} f^k(\hat{x}_0) &= \min_{y^k} \{(q^k)'y^k \mid Wy^k = h^k - L^k\hat{x}_0, y^k \geq 0\} \\ &= \max_{u^k} \{(h^k - L^k\hat{x}_0)'u^k \mid W'u^k \leq q^k\}. \end{aligned} \tag{2.18}$$

Solving each subproblem will generate a cut for the master problem, which could be an optimality cut or a feasibility cut, depending on whether the primal recourse program is feasible or not.

- **Case 1:** the primal problem is feasible, i.e. $f^k(\hat{x}_0)$ is finite. Then we have a dual optimal solution \hat{u}^k and a primal optimal solution \hat{y}^k . From (2.18), we have

$$f^k(\hat{x}_0) = (h^k - L^k\hat{x}_0)'\hat{u}^k, \tag{2.19}$$

while observe that for any value of x

$$\begin{aligned} f^k(x) &= \sup_{u^k} \{(h^k - L^kx)'u^k \mid W'u^k \leq q^k\} \\ &\geq (h^k - L^kx)'\hat{u}^k. \end{aligned}$$

Using the constraint $\theta^k - f^k(x) \geq 0$ in program (2.17), the above inequality yields:

$$\theta^k \geq (\hat{u}^k)'(h^k - L^kx).$$

By (2.19), we can rewrite the above equation as:

$$\theta^k \geq f^k(\hat{x}_0) - (\hat{u}^k)'L^k(x - \hat{x}_0)$$

This is the so-called *optimality cut* that the optimal pair $(\hat{x}, \hat{\theta})$ should satisfy.

- **Case 2:** the primal problem is infeasible, i.e. $f^k(\hat{x}_0) = +\infty$. This means \hat{x}_0 is indeed not feasible for all the constraints in (2.16), in which case the dual will be

unbounded. Then we can find a ray of unboundedness \hat{u}^k such that $W'\hat{u}^k \leq 0$ and $(h^k - L^k\hat{x}_0)'\hat{u}^k \geq 0$, since by multiplying by a positive number on the left-hand side of both equations, the constraints of dual problem are satisfied and the value of the objective function can grow to infinity. Meanwhile, given any feasible value x , some $y \geq 0$ exists such that $Wy = h^k - L^kx$. Hence, we have

$$(\hat{u}^k)'(h^k - L^kx) = (\hat{u}^k)'Wy \leq 0,$$

because $W'\hat{u}^k \leq 0$ and $y \geq 0$. Such inequality should hold for all feasible values of x ; observe that it is violated by \hat{x}_0 because $(h^k - L^k\hat{x}_0)'\hat{u}^k \geq 0$. Therefore we introduce the *feasibility cut*, cutting off the infeasible solution \hat{x}_0 , as

$$(\hat{u}^k)'(h^k - L^k\hat{x}) \leq 0.$$

After computing all r terms of the recourse function, that is solving all subproblems here, r cuts are generated, either optimality or feasibility cuts. Then, adding these cuts to the problem (2.17), the master problem can be replaced with:

$$\begin{aligned} \min_x \quad & c'x + \sum_{k=1}^r \theta^k \\ \text{s.t.} \quad & Ax = b \\ & \theta^k \geq f^k(\hat{x}_0) - (\hat{u}^k)'L^k(x - \hat{x}_0), \quad k \in K_1 \\ & (\hat{u}^k)'(h^k - L^k)\hat{x} \leq 0, \quad k \in K_2 \\ & x \geq 0, \quad K_1 \cup K_2 = \{1, \dots, r\}, \quad K_1 \cap K_2 = \emptyset, \end{aligned} \tag{2.20}$$

where K_1 and K_2 are the sets of optimality and feasibility cuts, respectively. The solution of the above master problem provides the next query point $(\hat{x}_1, \hat{\theta}_1)$, which can be substituted into the recourse subproblems. The computations can then all be repeated keeping all previously generated cuts.

The optimal solution to the master problem (2.20) provides a lower bound $\underline{\phi}$ to the original program (2.17). Observe that an upper bound of the optimal value of (2.17) can be generated as

$$\bar{\phi} = c'\hat{x}_0 + \sum_{k=1}^r f^k(\hat{x}_0),$$

if all subproblems are feasible, or it can be selected as $\bar{\phi} = +\infty$ otherwise. If

$$\bar{\phi} - \underline{\phi} \leq \varepsilon,$$

where ε is some small number to be chosen, the algorithm stops.

In all, the master problem conveys a query point to the subproblems which will generate cuts as feedback to the master problems and the solution of the master problem

provides an upper bound to the original program. Then solution of the master problem will provide a lower bound for the program and the next query point. When the lower bound and the upper bound are close enough, an optimum is achieved.

The following proposition can be found in Kall and Wallace [71]:

Proposition 2. *If the program (2.16) is solvable and its feasible set is bounded, the Benders decomposition method yields an optimal solution after finitely many steps.*

2.4.3 Interior Point Method and OOPS

Over the past two decades of linear programming research, interior point methods [127] have arisen as an alternative to the simplex method. Theoretical research and practical analysis have already shown their superior performance and efficiency over the simplex method for certain classes of problems. *Interior point methods* consistently require a small number of iterations to achieve the optimal solution, as well as fairly simple linear algebra: this makes the method well-suited to large scale problems. We know that the size of stochastic programming can easily grow with the number of realisations of random factors. This section provides a brief derivation of *interior point methods* and an introduction of OOPS [55, 57], a structure-exploiting parallel interior point method.

Primal-Dual Interior Point Methods

Interior point methods (IPM) provide a unified framework for optimisation algorithms for linear, quadratic and nonlinear programming. The explanation of their theoretical background can be found in [127]. In this section we present the algorithm for linear programming problems.

Consider a primal-dual pair of a linear program in a standard form as presented in (2.12) and (2.13), which is restated here for convenience:

$$\begin{array}{ll}
 \min & c'x \\
 (Primal) \quad s.t. & Ax = b \\
 & x \geq 0,
 \end{array}
 \qquad
 \begin{array}{ll}
 \max & b'u \\
 (Dual) \quad s.t. & A'u + s = c, \\
 & s \geq 0,
 \end{array}
 \tag{2.21}$$

where $c, x \in \mathbb{R}^n$, $b, u \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, and $m \leq n$, $s \in \mathbb{R}^n$ are so-called slack variables. In this section we will consider the dual form in (2.21). Without loss of generality, we can assume that A has full row rank, $rank(A) = m$. Specialising the Karush-Kuhn-Tucker (KKT) conditions for the above primal-dual pair, the vector $\hat{x} \in \mathbb{R}^n$ is a solution to the primal problem in (2.21) if and only if there exist vectors $\hat{u} \in \mathbb{R}^m$ and $\hat{s} \in \mathbb{R}^n$

for which the following conditions hold for $(x, u, s) = (\hat{x}, \hat{u}, \hat{s})$:

$$Ax = b, \quad (2.22a)$$

$$A'u + s = c, \quad (2.22b)$$

$$x_i s_i = 0, \quad i = 1, 2, \dots, n, \quad (2.22c)$$

$$(x, s) \geq 0. \quad (2.22d)$$

(2.22a), (2.22b) and (2.22d) imply primal and dual feasibility. Condition (2.22c) implies that for each $i = 1, \dots, n$, either x_i or s_i must be *zero*, which are called complementarity conditions, i.e. zeros in x and s appear in complementary positions. By multiplying with x' on both sides of (2.22b) and rearranging we obtain

$$c'x - b'u = x's,$$

implying the equivalence of complementarity and the duality gap for feasible points, which is *zero* at optimality as shown previously.

The vector $(\hat{u}, \hat{s}) \in \mathbb{R}^m \times \mathbb{R}^n$ is a solution to the dual problem in (2.21) if and only if there exists a vector $\hat{x} \in \mathbb{R}^n$ such that the conditions (2.22) hold for $(x, u, s) = (\hat{x}, \hat{u}, \hat{s})$. In conclusion, \hat{x} is the primal solution and (\hat{u}, \hat{s}) is the dual solution if and only if $(\hat{x}, \hat{u}, \hat{s})$ is the solution to the system (2.22). $(\hat{x}, \hat{u}, \hat{s})$ is called a primal-dual solution.

We rewrite the KKT conditions as

$$F(x, u, s) = \begin{bmatrix} Ax - b \\ A'u + s - c \\ XSe \\ (x, s) \end{bmatrix} = 0, \quad (2.23)$$

$$(x, s) \geq 0,$$

where X and S are diagonal matrices, $X = \text{diag}(x_1, \dots, x_n)$ and $S = \text{diag}(s_1, \dots, s_n)$, $e = (1, 1, \dots, 1)' \in \mathbb{R}^n$. This is a nonlinear system. The optimal solution can be achieved through solving the KKT system (2.23) by Newton's method. The primal-dual interior-point methods apply the Newton's method to achieve the optimal solution.

The primal and dual feasible sets can be written as:

$$\mathcal{P} = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}, \quad \mathcal{D} = \{(u, s) \in \mathbb{R}^m \times \mathbb{R}^n : A'u + s = c, s \geq 0\}.$$

The strictly feasible sets, or the set of primal interior points and the set of dual interior points, are given as

$$\mathcal{P}^0 = \{x \in \mathbb{R}^n : Ax = b, x > 0\}, \quad \mathcal{D}^0 = \{(u, s) \in \mathbb{R}^m \times \mathbb{R}^n : A'u + s = c, s > 0\}.$$

Then the primal-dual feasible set is

$$\mathcal{F} = \mathcal{P} \times \mathcal{D} = \{(x, u, s) \mid Ax = b, A'u + s = c, (x, s) \geq 0\},$$

and accordingly, the set of primal-dual interior points (the primal-dual strictly feasible set) is

$$\mathcal{F}^0 = \mathcal{P}^0 \times \mathcal{D}^0 = \{(x, u, s) \mid Ax = b, A'u + s = c, (x, s) > 0\}.$$

Given a scalar $\tau > 0$, solving the system

$$Ax = b, \tag{2.24a}$$

$$A'u + s = c, \tag{2.24b}$$

$$x_i s_i = \tau, \quad i = 1, 2, \dots, n, \tag{2.24c}$$

$$(x, s) > 0. \tag{2.24d}$$

gives a point (x^τ, u^τ, s^τ) , a primal-dual strictly feasible point. Instead of the complementarity condition and nonnegative bounds condition, the pairwise products $x_i^\tau s_i^\tau$ are all equal to a positive number τ for all $i = 1, \dots, n$ and these two vectors are strictly positive. Varying the value of τ , points (x_τ, u_τ, s_τ) construct an arc, which is the so called *central path*

$$\mathcal{C} = \{(x^\tau, u^\tau, s^\tau) \mid \tau > 0\}.$$

Given the strictly feasible set $\mathcal{F}^0 \neq \emptyset$, the point (x^τ, u^τ, s^τ) is unique for each $\tau > 0$. If the central path converges while driving τ to approach 0, it must converge to a primal-dual solution, e.g. Fig 2.4, where the bold curve is the central path.

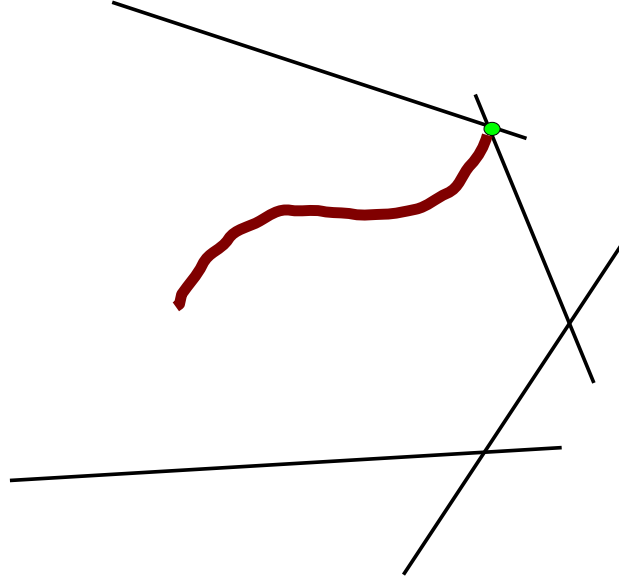


Figure 2.4: An example of a central path.

A Newton step $(\Delta x, \Delta u, \Delta s)$ towards the central path can be generated by solving

$$J(x, u, s) \begin{bmatrix} \Delta x \\ \Delta u \\ \Delta s \end{bmatrix} = -F(x, u, s) = - \begin{bmatrix} Ax - b \\ A'u + s - c \\ XSe - \sigma\tau e \end{bmatrix},$$

where J is the Jacobian of F , $\sigma \in [0, 1]$ is a *centering parameter* and τ is a *duality measure* defined by

$$\tau = \frac{1}{n} \sum_{i=1}^n x_i s_i = x' s / n.$$

The linear system is then

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A' & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta u \\ \Delta s \end{bmatrix} = \begin{bmatrix} \zeta^p \\ \zeta^d \\ \zeta^\mu \end{bmatrix}, \quad (2.25)$$

where

$$\zeta^p = b - Ax, \quad \zeta^d = c - A'u - s, \quad \zeta^\tau = \sigma\tau e - XSe.$$

The step $(\Delta x, \Delta u, \Delta s)$ is toward a point at which each pairwise product $x_i s_i$ is equal to $\sigma\tau$ in contrast to 0 required by (2.23). This step would be a standard Newton step of the KKT system (2.23) if $\sigma = 0$, while $\sigma = 1$ gives a centering direction that leads to a central point $(x^\tau, u^\tau, s^\tau) \in \mathcal{C}$ at which $x_i^\tau s_i^\tau = \tau$. Hence, trading off between reducing the duality gap τ and improving centrality is always needed, for example by taking an intermediate value of σ from the open interval $(0, 1)$.

”Since these steps are biased toward the interior of the nonnegative orthant defined by $(x, s) \geq 0$, it usually is possible to take longer steps along them than along the pure Newton steps for F before violating the positivity condition.” [127] p.7

By elimination of

$$\Delta s = X^{-1}(\zeta^\tau - S\Delta X) = -X^{-1}S\Delta X + X^{-1}\zeta^\tau,$$

from the second equation of (2.25), we get the symmetric indefinite augmented system of linear equations

$$\begin{bmatrix} -\Theta^{-1} & A' \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta u \end{bmatrix} = \begin{bmatrix} \zeta^d - X^{-1}\zeta^\tau \\ \zeta^p \end{bmatrix} \quad (2.26)$$

to be solved, where $\Theta = XS^{-1}$ is a diagonal matrix. In addition, the constraint matrix A in stochastic programming presents a block-angular structure, for instance Fig 2.3. And as a result, the whole augmented system (2.26) displays a block-sparse pattern for stochastic programming problems which can be exploited to make the solution process

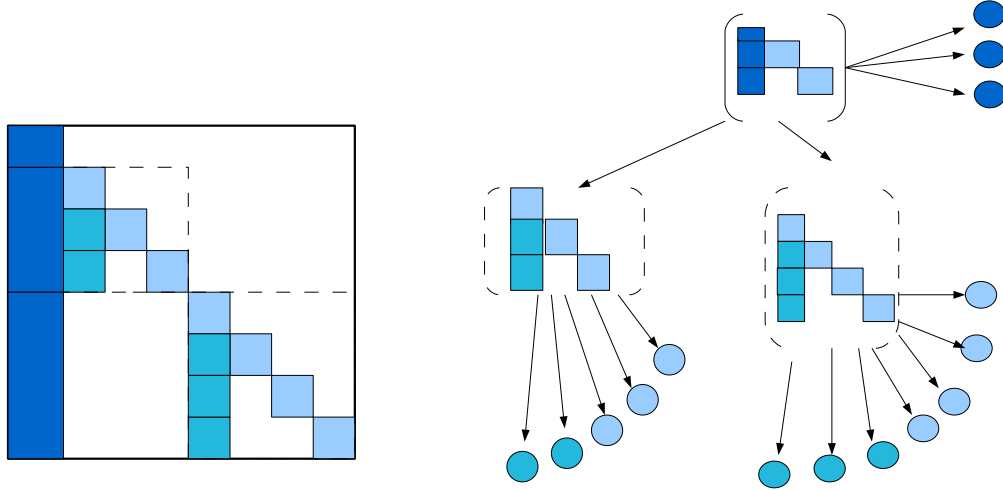


Figure 2.5: Nested block-structured constraint matrix with its tree representation

more efficiently.

Object-Oriented Parallel Solver

An Object-Oriented Parallel Solver (OOPS) [55, 57] is based on primal-dual interior-point method. This solver can exploit the nested block structure of the problem by representing the matrices defining the problem as a tree of structured matrices. A linear algebra kernel which can exploit special structure is used inside a primal-dual interior point solver targeted at convex optimization problems. It is able to solve large scale linear or quadratic problems efficiently, by exploiting the block structure of the matrices in the augmented system (2.26). The matrices with the block structure are composed of sub-matrices; furthermore, each sub-matrix could be a block-structured matrix itself, which shows a nested block structure. Such matrices appear in a number of mathematical programming problems. Stochastic programming is one of the class of problems that have such properties. More nesting of blocks is involved as the number of stages modelled is increased.

In OOPS, each matrix block is treated as an object. A linear algebra is designed to exploit the structure information of the matrix and to build up a representing tree with the blocks as nodes, i.e. the whole matrix is represented as the root, a child of a node is a sub-matrix of the matrix represented by this node and leaf nodes are elementary matrices that cannot be broken down further into smaller blocks. Figure 2.5 illustrates how to build up a representing tree of the multi-stage stochastic program shown in Figure 2.3. The structure information of each block is associated with the corresponding node.

Then, every primal-dual interior point iteration will be working on this tree. The leaf nodes are evaluated first. The evaluation of a matrix corresponding to a non-

leaf node can be distributed to a sequence of operations on its sub-blocks represented by its child nodes. The operations on those sub-blocks are independent, and can be processed in a parallel manner. What is worth mentioning is that the exact sequence of operations at a given node depends on the structure of the corresponding matrix; however, the structure information of the parent and child node is not related to each other. Therefore, instead of solving the whole large augmented system as one, it is decomposed to a sequence of operations on several smaller blocks according to the structure of the matrix and those blocks may be decomposed further.

Both Benders decomposition and OOPS work in a fashion that breaks the large problem into several small ones and they both suit most stochastic linear programs.

Using Benders decomposition, it is even not necessary to write and input the large whole deterministic equivalent to the algorithm, namely one can just write the smaller master problem and subproblems to save memory. However, the increase of the stage of stochastic programs may lead to significant complexity in the decomposition algorithm. Different algorithms will be needed for linear problems (LP) and nonlinear problems (NLP). As Benders decomposition relies on the diagonal block structure (nonzero's only appear in the diagonal blocks besides a border of columns), the challenge will come if constraints linking the blocks together appear, as we will see in Chapter 3.

Since it is embedded in an IPM solver, OOPS is designed to deal with convex problems, both LP and NLP. One can use OOPS just by inputting the model and the structure of the model which is normally known through the modelling process. The algorithm does not require any modification to solve models with more stages. Furthermore, the block structure that OOPS can deal with allows for any combination of nested block structures, including the diagonal block structure.

2.5 Exogenous and Endogenous Problems

All stochastic programs and related recourse problems discussed above assume that the probability distributions of the random parameters are all independent of the decision variables. Recall stochastic program (2.2)

$$\begin{aligned} \text{"min"} \quad & f_0(x, \bar{\xi}) \\ \text{s.t.} \quad & g_i(x, \bar{\xi}) \leq 0, \quad i = 1, \dots, m \\ & x \in X \subset \mathbb{R}^n, \end{aligned}$$

where $\bar{\xi}$ is a random vector varying over a set Ω with probability measure \mathcal{P} , or say in the probability space $(\Omega, \mathcal{F}, \mathcal{P})$. The probability measure \mathcal{P} in most cases is independent of the decision variables x_j , $j = 1, \dots, n$, and these kinds of problems are referred to as exogenous problems. Different from them are the so called endogenous problems, which deal with some important decision problems in which, through the decision process, an endogenous randomness depending on decision variables occurs. In

the classical form of stochastic programming, the decision variables and uncertainties are independent, see [13, 37, 71, 106]. The stochastic programs modelled with the so called endogenous uncertainty, in which decision variables can influence the uncertainties, were first addressed by Pflug [103].

As an example, consider the gas field development planning problem presented in [50]. There is a set of gas reservoirs available for production, with unknown sizes and qualities of the reserves. This uncertainty is resolved only when a facility is installed at that field. Thus, the timing of uncertainties to be resolved depends on the investment decisions. It is important to consider the potential of obtaining valuable information as a result of the investment, besides considering the capital expenditures and associated revenues.

Ignoring the dependence by formulating a simpler model is a way to deal with this complicated problem. However, it is not a satisfactory approach. Embedding this dependence in the model may cause various technical difficulties, e.g. if $f_0(x, \bar{\xi})$ is convex in x for each realisation ξ of $\bar{\xi}$, then it is also convex for the problem with independent probability measure, whereas the convexity property may be lost for \mathcal{P}_x (i.e. when the probability distribution \mathcal{P} is a function of the decision variable x).

Endogenous problems are generally classified in two types:

- the probability distribution is of a known type and only its parameters depend on decisions;
- there is a fixed finite set of probability distributions and the one to be chosen depends on decisions. Such dependence may be modeled by Boolean variables and decisions may be partly related to the choice of probability distribution from the given set.

A hybrid mixed-integer disjunctive programming formulation for stochastic programming corresponding to endogenous problems is presented in [50] with a branch and bound algorithm, while a decomposition algorithm for a model with boolean variables is proposed in [68].

While most work in non-standard problems is about the decision-dependent probability distributions, in the workforce planning problem in Chapter 4 we are dealing with uncertain elements in the model (e.g. service demand) depending on decisions. This is due to the fact that the labour allocation strategies will determine the output of the service, which could have an impact on the reputation of the institution. While customers make choices of the companies to deliver service according to their reputations, demands of the institution could be further influenced.

The reader interested in further discussions of non-standard problems is referred to [5, 36, 50, 68] and the references therein.

2.6 Application Areas

Stochastic programming has been used to solve problems in the fields where uncertainty is an important factor. With the aid of fast-growing computational capability, stochastic programming methodology becomes applicable even in complex problems and new areas. Here we list a few categories of problems in which this technique is considered to be reliable and in which it is well developed.

- Finance, including fund management, risk management, derivative product pricing. Asset-Liability Management [132] is a well-known example of a financial problem dealt with by stochastic programming, where both asset returns and liabilities are uncertain. This problem will be discussed in detail in Chapter 3. Further developments in this area consider the strategy-making of pension schemes, i.e. how to set the minimum payment by customers in order to maintain a sustainable pension fund. Stochastic programming can also be used in market risk and credit risk management in the financial market [4, 54]. Some work has already shown how to price derivative products by stochastic programming [22, 23]. The principle is using the replication rule, which means the return of such product can be replicated by a set of simple financial products whose prices are already known. By the arbitrage free assumption, the price of this product should be equal to this set of assets and such a set is determined by minimising the difference in returns between the product and the set of assets.
- Management. Production planning is one of the classical application areas of stochastic programming, which seeks to maximise profit when resources are limited [3, 40]. In such problems, demand, as well as the prices of resources and the sale price of products, is normally random. It is important to take these uncertainties into account to avoid unsatisfied demand, lack of resources or drop in profit. Some manufactures may also look for opportunities to expand their capacities in the situation in which both expanding cost and induced profit cannot be predicted, considering how these can be organised efficiently so that high revenue with low cost is attainable in the long term. On the other hand, human resource allocation can also be planned by stochastic programming, to find an optimal strategy such that all the work requested is completed on time with lowest managerial cost, while the available workforce or work due may be uncertain. The problem dealt with in Chapter 4 is concerned with the human factor in operational risk management.
- Energy industry, e.g. the electricity and gas markets. A typical application of stochastic programming is electricity generation scheduling and planning. There are three broad classes of problems having received much attention [128, 29], i.e. unit-commitment problems [120, 119], hydro-thermal scheduling [20, 95] and capacity expansion [2, 85]. In the most general form, all of these problems can be

modelled as multi-stage stochastic programming problems. In addition, stochastic equilibrium modelling for electricity markets is also a new development of stochastic programming [94]. In gas markets, natural gas is produced in production fields and transported in pipelines by using compressors such that molecules flow from the originating end of a pipeline with a high pressure towards the another end. The planning problems for the producers are uncertain, due to the price uncertainty and flexibility of pipeline systems and contracts [62]. Stochastic mixed integer programming is used to model supply chain optimisation for coordination of production, transportation, storage and contract management.

- Logistics. The efficient design of the supply chain is crucial to manufacturing, which involves a network organisation of suppliers, manufacturing plants, warehouses, and distribution channels, so as to acquire raw materials, convert these raw materials to finished products, and distribute these products to customers. The costs and demands are both difficult to predict. The production and distribution have to be optimised based on these uncertainties, which can be handled by stochastic programming, [115].
- Telecommunications. The demand of network capacity is scarce resource due to the growing requirement of higher bandwidths. In the meantime, such a demand requirement presents an opportunity for significant revenue growth, under the condition of enough resources are available. Because of the competitive telecommunication market customers can choose the network provider they prefer. Therefore, the network provider has to plan bandwidth allocation through network links carefully so as to maximise the potential number of requests served by the network. The function of the network and demand are both random in such problems. Stochastic programming was used to determine the optimal bandwidth such that the maximal revenue is achieved, [116].

Chapter 3

Measuring and Managing Market Risk by Stochastic Dominance

Stochastic programming has been used to aid portfolio decision making for over twenty years. The control of market risk is one of the most important aspects of this application. In this chapter, we will work on multi-stage Asset-Liability Management modelling and explore efficient risk control methods. Furthermore, OOPS, an object-oriented parallel solver, is used to solve the resulting optimisation problems and its sound performance is demonstrated by the computational results compared with another solver at the end of this chapter. The original content of most parts of this chapter has already appeared in [129], coauthored with Jacek Gondzio and Andreas Grothey.

3.1 Asset-Liability Management

ALM models assist financial institutions in decision making on asset allocations considering full use of the fund and resources available. The model aims to maximise the overall revenue, sometimes as well as revenue at intermediate stages, with restrictions on risk. Risk in ALM problems is present in two aspects: a possible loss of investment and missing the ability to meet liabilities. The returns of assets and the liabilities are both uncertain. It is essential in ALM modelling to deal with uncertainties as well as with risks. The stochastic programming approach is naturally applicable to problems which involve uncertainties; an approach (with stochastic dominance) to deal with risk management is discussed in the next section.

3.1.1 Literature Review of Asset-Liability Management Modelling

The asset and liability management problem is essential to insurance companies and banks in which the business involves large amount of liquidity. By liquidity, we mean that the financial sector has to satisfy liability while pursuing profit. The liability

mainly comes from savers' deposits for banks, and from benefits paid to insurees for insurance companies. A collection of scientific papers about ALM can be found in [132].

In [93], ALM modelling is reviewed from a financial engineering point of view. The authors discuss 3 issues in ALM modelling: modelling, scenario generation and solution techniques. About modelling, the authors emphasise 7 factors that should be considered in modelling:

1. multiperiodity, corresponding to the dynamic nature of the problem including return over time, cash flows, rebalance, etc;
2. adequate treatment of uncertainty, which is involved in certain parameters, e.g. returns, cash flows. These uncertainties result in the liquidity risk;
3. risk management, ability to account for the decision maker's risk-bearing attitude;
4. consideration of transaction costs, implying the expenses incurred in transactions, e.g. commissions;
5. integration of asset and liabilities, addressing the financial planning process;
6. understandability, that the model can be explained to fund managers and other users;
7. other factors, e.g. legal, institutional, policy constraints.

A major difficulty in modelling the ALM problem consists in risk management. In 1986, Kusy and Ziemba [80] proposed a model for a bank that maximises revenues in the objective while satisfying several categories of constraints composed of legal, policy, liquidity, cash flow and budget constraints which make sure that deposit liability is satisfied as much as possible. The main methodology adopted was stochastic modelling with simple recourse and linear programming, and it was practical even for large banks under the computational limit at that time. However, the market risk of assets was not well defined or measured.

The Markowitz risk-averse paradigm [86] is also considered to model ALM problem and optimise multiple objectives: maximise the return and minimise the associated risk, e.g. [107]. Several techniques can also be used to model the problem in stochastic form, e.g. chance-constraints [18], dynamic programming [26], and sequential decision [126]. A successful example of optimisation-based ALM modelling which took risk management issues into account was the Russell-Yasuda Kasai model for a Japanese insurance company by the Frank Russell consulting company, which used multi-stage stochastic programming [16, 17]. This dynamic stochastic model took into account multiple accounts, regulatory rules and liabilities to enable the managing of complex issues arising in the Yasuda Fire and Marine Insurance company. Expected shortfall, i.e. the expected amount by which the goals were not achieved, was applied to measure risk more accurately than the calculation of expected penalties and it was easy to handle

in the solution process. Moreover, the model proved to be easy for decision-makers to understand. The implementation results showed the advantages of the Russell-Yasuda model over the mean-variance model in multi-period and multi-account problems.

Conditional Value-at-Risk was applied to an ALM problem for modelling pension funds in [14]. The ratio of assets to liabilities is referred to as the funding ratio of the pension fund and a target funding ratio is incorporated into the model, which can be different from stage to stage. How the target funding ratio is satisfied by asset value is constrained by Conditional Value-at-Risk. In addition to portfolio strategies, the contribution rate is also decided in the model. The objective of the model is to minimise the contribution from employees and employers as well as the loans involved. Instead of using a scenario tree or event tree, the authors introduced bundles, which grouped some sample paths at the time point until where the paths from the starting point are the same. The numbers of asset units are the same in this bundle, i.e. they correspond to the same strategy. This can ease the computational burden due to a decrease in the number of scenarios while maintaining the stochastic nature of the problem.

The concept of stochastic dominance dates from the work of Karamata in 1932 where he proved a theorem which is quite similar to the second-order stochastic dominance (see [83] for a survey). Subsequently, similar concepts have been applied in statistics. The application to decision theory began about half a century ago. However, only after the end of 1960s was the theory of stochastic dominance developed, as well as its theoretical and empirical extensions in economics and finance. Stochastic dominance as a risk control tool has recently gained substantial interest from the research community. It has several attractive features but two of them are particularly important: stochastic dominance is consistent with utility functions and it considers the whole probability distribution and provides a partial order of random variables.

To the best of our knowledge, stochastic dominance has not yet been applied in the ALM context and in this work we demonstrate how this can be done. Further, we develop a chance constraint from relaxed interval second-order stochastic dominance and show that it is an intermediate dominance constraint between first-order and second-order with a discrete probability distributions. By combining second-order stochastic dominance and relaxed interval second-order stochastic dominance, the model can help generate portfolio strategies with better management of risk and better control of underfunding. We illustrate this issue with a small example analysed in Section 3.4.1 and a backtest in Section 3.4.2.

As the second issue, scenario generation is also critical to the programming. This includes generating scenarios for three uncertain parameters: economic factor (e.g. interest rate), return of assets, and liability to be satisfied. Scenarios generated constitute the universe of possible outcomes. The problem requires parameter estimation of those random variables, which can be done by several techniques, such as maximum likelihood, generalised method of moments, simulated moment estimations and integrated

parameter estimation, to mention just a few of them.

The third issue concerns with the solution technique. The problem can be solved directly as a linear or nonlinear program. Since traditional stochastic programming gives a tree structure of constraints, decomposition is considered efficient by taking advantage of such special structure. The model named CALM was proposed in SMPS format by Consigli and Dempster in [21] using dynamic recourse programming and generating a large scale deterministic equivalent problem. They applied the nested Benders decomposition method to solve it, which was shown at the time to be more efficient than the general simplex method or interior point method, in terms of speed and memory requirement.

Three optimisation technologies that can be applied to the portfolio selection problem i.e. stochastic control, stochastic programming and Monte Carlo simulation, are also reviewed and compared in [92] by Mulvey. When a solution is achieved, stochastic control represents an ideal framework that is easy to understand and implement, but it is limited to small problems and hard to solve in analytic form. Stochastic programming provides a general purpose decision model; however the size of the program may grow quickly with the length of the planning horizon and the number of scenarios. Rather than seeking an optimal strategy, Monte Carlo aims to find the best set of parameters for a prespecified policy rule. It is the simplest approach among these three and is easy to implement.

Risk constraints may link variables which are associated with different nodes at the same stage in the event tree. Adding such constraints to the stochastic programming problem destroys the usual tree-structure of the problem and prevents the use of Benders decomposition and progressive hedging [108]. However, such problem still presents a special structure, which is shown in Section 3.4.3. We convey such structure of ALM models to a specialised structure-exploiting parallel interior point solver OOPS [57] which takes advantage of such information in the solution process and can effectively deal with complicated ALM problems which contain special constraint structure. The analysis of computational results confirms that, by exploiting the structure, OOPS outperforms the commercial optimisation solver CPLEX 10.0 on these problems.

3.1.2 Multi-Stage ALM Modelling

Generally, ALM modelling follows the mean-risk methodology. Besides the return and risk control, the ALM model also has the following features:

1. Transaction costs; each transaction will be charged at a certain percentage of total transaction value, and different transaction costs may apply to purchases and sales;
2. Cash balance; liabilities should be paid to clients, meanwhile there is an inflow in terms of deposits or premiums; the model should make sure the outflow and inflow match;

3. Inventories of assets and cash, which are essential in a dynamical systems of 2- or even a multiple - stage problems;
4. Legal and policy constraints aligned with the financial sector's requirements.

We only consider the first three features in ALM model.

It is important for the decision makers to rebalance the portfolio during the investment period as they may wish to adjust the asset allocations according to updated information on the market. The strategy which is currently optimal may not be optimal any more as the situation changes. Taking this into account, the problem is a multi-period problem and at the beginning of each period, new decisions are made.

We denote the time horizon by T . At each time stage t , $t = 0, \dots, T$, a decision is made on the units of each asset to be invested in and amount of cash held, based on the state of the total wealth and the forecast of prospective performances of the assets at that particular time. When the random factors follow discrete distributions, the resulting decision process can be captured by an event tree, as shown in Figure 3.1. Each node is labelled with (i, j) denoting node j at stage i . Each node represents a possible future event. Asset returns, liabilities and cash deposits are subject to uncertain future evolution. Meanwhile, the asset rebalancing is done after knowing which value the asset returns and liabilities take at each node.

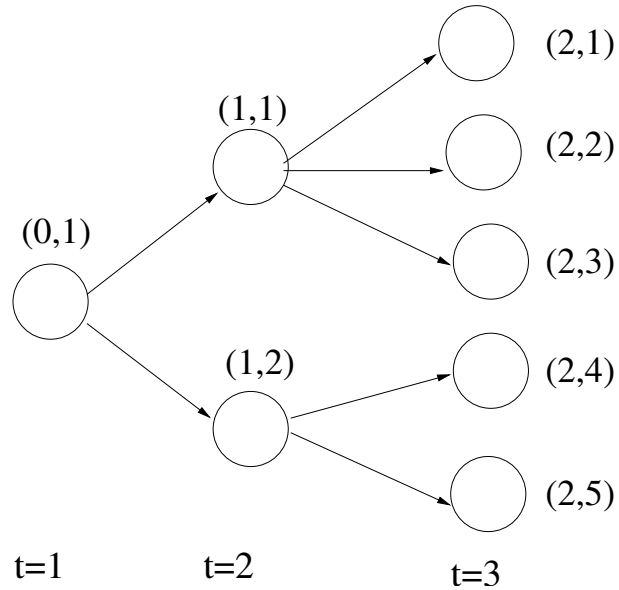


Figure 3.1: An example of event tree describing different return states of nature.

The notation of the model is given first:

Parameters:

W_i : price of asset i ;

G : total initial wealth;

λ : the penalty coefficient of underfunding;

γ : the transaction fee, which is proportional to trading volume (assumed to be equal

for purchases and sales);

β : upper bound on acceptable risk;

ψ : funding ratio, showing the percentage of liabilities to be satisfied;

Random data:

$R_{i,j}^t$: the return of asset i in node j at stage t ;

$R_{c,j}^t$: the interest rate in node j at stage t ;

A_0, A_j^t : the outflow of resources, e.g. liabilities;

D_0, D_j^t : the inflow of resources, e.g. contributions;

π : the joint probability distribution of above uncertain factors, i.e. $\pi_{j_t}^t$ is the probability of node j_t occurring;

Decision variables:

$xh_{i,j}^t$: units of asset i held in node j at the beginning of stage t ;

$xs_{i,j}^t$: units of asset i sold in node j at the beginning of stage t ;

$xb_{i,j}^t$: units of asset i bought in node j at the beginning of stage t ;

c_j^t : units of cash held in node j at stage t ;

b_j^T : the amount of underfunding in node j at the terminal stage that cannot be satisfied;

Indexes and sets:

t : the stage index, with $t = 1, \dots, T$;

i : the asset index, with $i \in I = \{1, \dots, m\}$;

n_t : the number of nodes at stage t ;

j_t : the node index, with $j_t \in N_t = \{1, \dots, n_t\}$, $t = 1, \dots, T$;

$a_{(t,j_t)}$: the ancestor of node $(t, j(t))$;

In this work we will use formulation (1.3), i.e. maximise expected return with acceptable risk. Then the multi-stage ALM problem concerning the investment strategy can be represented as:

$$\max \sum_{j_T \in N_T} \pi_{j_T}^T \left(\sum_{i \in I} (1 - \gamma) W_i x h_{i,j_T}^T + c_{j_T}^T - \lambda b_{j_T}^T \right) \quad (3.1a)$$

$$s.t. \quad (1 + \gamma) \sum_{i \in I} W_i x h_{i,0}^0 + c_0 = G - A_0 + D_0 \quad (3.1b)$$

$$(1 + \gamma) \sum_{i \in I} W_i x b_{i,j_t}^t + c_{j_t}^t = (1 - \gamma) \sum_{i \in I} W_i x s_{i,j_t}^t + (1 + R_{c,j_t}^t) c_{a(t,j_t)}^{t-1} - A_{j_t}^t + D_{j_t}^t, \quad (3.1c)$$

$$j_t = 1, \dots, n_t, \quad t = 1, \dots, T,$$

$$(1 + R_{i,j_t}^t) x h_{i,a(t,j_t)}^{t-1} + x b_{i,j_t}^t - x s_{i,j_t}^t = x h_{i,j_t}^t, \quad (3.1d)$$

$$i \in I, \quad j_t = 1, \dots, n_t, \quad t = 1, \dots, T,$$

$$\sum_{i \in I} (1 - \gamma) W_i x h_{i,j_T}^T + c_{j_T}^T + b_{j_T}^T \geq \psi A_{j_T}^T, \quad (3.1e)$$

$$j_t = 1, \dots, n_t, \quad t = 1, \dots, T,$$

$$\phi(x h_{i,j_t}^t, c_{j_t}^t) \leq \beta, \quad j_t = 1, \dots, n_t, \quad t = 1, \dots, T, \quad (3.1f)$$

$$xh_{i,j_t}^t \geq 0, \quad xs_{i,j_t}^t \geq 0, \quad xb_{j_t}^t \geq 0, \quad b_{j_T}^T \geq 0, \quad xh_{i,j_t}^t, \quad xs_{i,j_t}^t, \quad xb_{i,j_t}^t \in \mathbb{R}^m$$

$$i \in I = \{1, \dots, m\}, \quad j_t \in N_t = \{1, \dots, n_t\}, \quad t = 1, \dots, T,$$

where $\phi(\cdot)$ gives the risk associated with position (xh, c) .

The decision maker does not seek a strategy to strictly satisfy the liability at the horizon of the problem, but penalises the underfunding. The objective (3.1a) aims to maximize the final wealth of the fund taking into account the penalties of underfunding. (3.1b) balances the initial wealth at the first stage while (3.1c) are cash balances for the following stages, both taking into account transaction cost, proportional to the total trade volume. The inventories of each asset at each stage are captured in (3.1d). (3.1e) defines the underfunding level b_j at the terminal stage. Risk control is expressed in (3.1f) with the risk measure function $\phi(\cdot)$ and the maximum acceptable level of risk β . This constraint will be discussed in more detail in the following section. If the risk constraint is linear, the model (3.1) is a linear program.

Risk control in an ALM problem involves many aspects. Two of the most important are overall performance and underfunding. The overall performance is analyzed considering all possible outcomes of the portfolio, e.g. variance. We will use stochastic dominance to control the risk of overall performance and discuss the modelling issues involved in Section 3.2. Underfunding concerns the possibility of unsatisfied liabilities. To avoid underfunding completely is expensive to implement and in many situations impossible. We will control underfunding through stochastic dominance constraints discussed in Section 3.2.5.

3.2 Stochastic Dominance

Stochastic dominance has been considered to be a reference to other risk measures by Ogryczak and Ruszczyński in [99]. Below we describe how it can be incorporated into our ALM model. First we briefly recall the definitions of stochastic dominance following closely the exposition in [99].

3.2.1 Definitions of Stochastic Dominance

Given a random variable ω , we consider the first performance function, which is actually the probability distribution function, as:

$$F_\omega^1(\eta) = P(\omega \leq \eta), \quad \eta \in \mathbb{R}.$$

Definition 2. *A random variable Y dominates L by first-order stochastic dominance (FSD) if and only if:*

$$F_Y^1(\eta) \leq F_L^1(\eta), \quad \forall \eta \in \mathbb{R}, \quad (3.2)$$

denoted as

$$Y \succeq_1 L.$$

Next, we define the second performance function as:

$$F_{\omega}^2(\eta) = \int_{-\infty}^{\eta} F_{\omega}^1(\zeta) d\zeta, \quad \forall \eta \in \mathbb{R}. \quad (3.3)$$

This function is continuous, convex, nonnegative, and nondecreasing.

Definition 3. A random variable Y dominates L by second-order stochastic dominance (SSD) if and only if:

$$F_Y^2(\eta) \leq F_L^2(\eta), \quad \forall \eta \in \mathbb{R}, \quad (3.4)$$

denoted as

$$Y \succeq_2 L.$$

Hence, if y and l are returns of two portfolio strategies satisfying (3.2) (or (3.4)), then Y dominates L and Y is preferable. Furthermore, we can define higher order performance functions recursively:

$$F_{\omega}^k(\eta) = \int_{-\infty}^{\eta} F_{\omega}^{k-1}(\zeta) d\zeta, \quad \forall \eta \in \mathbb{R}.$$

Definition 4. We say that a random variable Y dominates another random variable L by k -th order stochastic dominance if and only if

$$F_Y^k(\eta) \leq F_L^k(\eta), \quad \forall \eta \in \mathbb{R},$$

denoted as

$$Y \succeq_k L.$$

It is obvious that the lower order dominance relations guarantee the dominance of higher orders, see [99, 113]. A small example of this can be found in Section 3.2.5.

Stochastic dominance has been widely used in decision theory [43], economics [60] and finance [111] due to several advantages it offers for comparing random variables. It takes the entire probability distribution of random variables into account and provides partial orderings of those variables. Instead of setting fixed thresholds in portfolio selection model, it enables the use of random reference outcomes. In addition, it is consistent with utility theory.

Utility measures a degree of satisfaction. The value of a portfolio depends only on itself and is equal for every investor; the utility, however, is dependent on the particular circumstances of the person making the estimate. Investors seek to maximise their utilities. In general, utility functions are nondecreasing, which means most people prefer more fortune to less. It is known that $X \succeq_1 Y$ if and only if $E[U(X)] \geq E[U(Y)]$ for every nondecreasing utility function U for which these expected values are finite. And, $X \succeq_2 Y$ if and only if $E[U(X)] \geq E[U(Y)]$ for every nondecreasing and concave utility function U for which these expected values are finite. A nondecreasing and concave utility function reflects the fact that the investor prefers more fortune but the

speed of increase in satisfaction decreases. A survey of stochastic dominance and utility theory can be found in [83]. The proof of the consistency of stochastic dominance with utility theory and further analysis in this topic can be found in [61], with a modified version in [121]. Generally, a reasonable risk averse investor has a nondecreasing and concave utility function. We will incorporate SSD into the ALM models also because of its computational advantage, as we will show later, while FSD leads to a mixed integer formulation which can be found in [53, 96].

3.2.2 Literature Review of Stochastic Dominance

In [99], Ogryczak and Ruszczyński pointed out the importance of mean-risk models being consistent with stochastic dominance relations due to the fundamental role of the stochastic dominance concept in decision theory. This consistence is defined as follows in the same paper:

Definition 5. *We say the mean-risk model (E_X, r_X) , where E_X is the expected value and r_X is the risk, is consistent with second-order stochastic dominance if the following relation holds:*

$$X \succeq_2 Y \quad \Rightarrow \quad E_X \geq E_Y \quad r_X \leq r_Y. \quad (3.5)$$

The first inequality on the right-hand side is guaranteed by second-order stochastic dominance: $X \succeq_2 Y \Rightarrow E_X \geq E_Y$. The inequality representing the risk term, however, is not satisfied for some risk measures, like variance for instance. It has been shown in [99] that, $(E_X, CVaR_X)$ satisfies (3.5), while the mean-risk model with mean absolute deviation (MAD) as the risk measure is not consistent with second-order stochastic dominance. But still,

$$X \succeq_2 Y \quad \Rightarrow \quad E_X - MAD_X \geq E_Y - MAD_Y.$$

The efficient frontier, where all optimal portfolios are located, is constructed by these two mean-risk models in [113]. We will give our proof of monotonicity of CVaR with second-order stochastic dominance in a later section.

An application of the first-order stochastic dominance in the stochastic programming context leads to a non-convex mixed integer programming formulation. In contrast, second-order stochastic dominance can be incorporated in the form of a set of linearised constraints [30] which makes it a more attractive option. In a series of papers Dentcheva and Ruszczyński analysed several aspects of the use of stochastic dominance.

In [30], it is proved that the second-order stochastic dominance constraints construct a convex and closed set, i.e. for any $Y \in L^1(\Omega, \mathcal{F}, \mathcal{P})$ which is a probability space, $A_2(Y) = \{X \in L^1(\Omega, \mathcal{F}, \mathcal{P}) : X \succeq_2 Y\}$ is convex and closed. Furthermore, the optimality and duality conditions in general linear problems were also proved in this paper. Similar properties hold for the situation with multiple dominance constraints and for higher order dominance. An alternative approach to mean-risk portfolio models

was also provided by using stochastic dominance. Moreover, it was shown that the Lagrange multiplier associated with the dominance constraint can be identified with a certain concave and nondecreasing utility function. An application to static portfolio selection with a utility function constructed based on the methodology of Lagrange can be found in [32]. The objective is to maximise the expected return of the portfolio, and the stochastic dominance constraint is used to guarantee that the portfolio will not underperform certain criteria. The use of such a model eliminates the need to decide on a risk measure, as well as the need to choose the weight of the risk measure. Otherwise one needs to consider the entire family of risk measures and the arbitrary nature of risk measures makes it difficult to make a comparison. In addition, it is difficult to elicit utility functions from decision makers. When there is a group of decision-makers who have to come to a consensus, the situation becomes even more complicated. The requirement in the model with stochastic dominance, however, is a reasonable benchmark random outcome, e.g. market index, and this saves a lot of work.

Optimisation problems involving nonlinear stochastic dominance constraints, where stochastic dominance is used to compare nonlinear functions of random factors, were considered in [31]. This paper focussed on second-order stochastic dominance. Sufficient and necessary optimality conditions were also analysed. Their newly developed optimality and duality theory for this special class of problems also allows the creation of a decomposition approach to the problem, which they illustrated with a portfolio example.

Linear stochastic programming problems with first-order stochastic dominance are non-convex due to the non-convexity of the feasible set. The use of first-order stochastic dominance constraints, which are equivalent to a continuum of probabilistic constraints, introduces serious complications into the optimisation models and makes their solution difficult. Relaxations of these constraints were analysed in [96]. Based on second-order dominance constraints, Noyan, et al. formulated a linear first-order dominance constraints optimisation model as a mixed 0-1 programming problem with multiple knapsack constraints. The authors also introduced interval second-order stochastic dominance which is equivalent to first-order stochastic dominance and generated a mixed integer problem based on this dominance relation in [96]. Furthermore, in [97], a cutting plane method was employed to solve the program with first order dominance constraints. Three heuristic algorithms were generated to construct feasible solutions to these models, based on second-order dominance, variable fixing, and Conditional Value-at-Risk.

The conditions of stability and optimality of first-order stochastic dominance with respect to general perturbation of the underlying probability measures were given in [28]. The upper bound and lower bound of optimal values were also generated under certain conditions which illustrated the sensitivity of the optimal value function when the random variables in the problem are subject to perturbations.

Roman, et al. [110], proposed a multi-objective portfolio selection model with second-order stochastic dominance constraints to track or overperform a reference point, while Fábíán et al. [42] developed an efficient method to solve this model based on a cutting-plane scheme. The cutting-plane method transforms the model from a uniform discrete probability distribution through conditional value-at-risk into integrated chance constraints (ICCs). Then using the Künsi-Bay cutting plane method on ICCs, the same algorithm can be applied to the second-order stochastic model, which is just a special case of ICCs. The numerical results demonstrated the effectiveness of the solution algorithm.

The application of stochastic dominance in dispersed energy planning and decision problems, where the decision variables are integer, has been illustrated in [51, 52, 53] in the form of a mixed integer problem, including both first-order and second-order stochastic dominances. Stability and structural properties of the integer problems with dominance constraints were analysed in these papers. The authors applied a branch and bound decomposition algorithm to solve the problems.

The use of multivariate stochastic dominance to measure multiple random variables jointly was discussed in [33]. Again, the necessary and sufficient conditions for optimality and duality relations were developed for problems with these constraints in both convex and non-convex cases.

3.2.3 Monotonicity of CVaR with SSD

It has been proved that the mean-risk model (E_X, r_X) with r_X as the Conditional Value-at-Risk is consistent with SSD. We could not find the original paper with the proof and our proof is illustrated in the follows.

Given two portfolios with return as X and Y , for which the probability distribution functions and probability density functions are F_X , F_Y , and f_X , f_Y , respectively. And they both have limited expected values. For any probability $\beta \in [0, 1]$, let β -VaRs and β -CVaRs for the two portfolios be denoted as $\alpha(\beta)$ and $\phi(\beta)$, respectively. Portfolio X dominates portfolio Y by SSD is defined as

$$\int_{-\infty}^{\eta} F_X(X \leq \xi) d\xi \leq \int_{-\infty}^{\eta} F_Y(Y \leq \xi) d\xi \quad \forall \eta \in \mathbb{R},$$

and the inequality holds strictly for at least one point of η .

Theorem 1. *The mean-risk model $(E_X, CVaR_X)$, where E_X is the expected value and $CVaR_X$ is the risk measure, is consistent with second-order stochastic dominance.*

Proof. From Definition 5, to prove the consistence of the mean-risk model is to prove the following:

$$X \succeq_2 Y \quad \Rightarrow \quad E_X \geq E_Y \quad CVaR_X \leq CVaR_Y. \quad (3.6)$$

Since the first inequality on the right-hand side is guaranteed by second-order stochastic dominance, we only need to prove the inequality for CVaR.

We write CVaRs explicitly as follows:

$$\phi_X(\beta) = \frac{1}{1-\beta} \int_{-\infty}^{\alpha_X(\beta)} \xi f_X(\xi) d\xi,$$

$$\phi_Y(\beta) = \frac{1}{1-\beta} \int_{-\infty}^{\alpha_Y(\beta)} \xi f_Y(\xi) d\xi.$$

In the integrations of CVaRs, both VaRs, i.e. $\alpha_X(\beta)$ and $\alpha_Y(\beta)$, are real numbers.

Case 1: $\alpha_X \leq \alpha_Y$.

In this case,

$$F_X(\alpha_Y) \geq F_X(\alpha_X) = \beta.$$

Then,

$$\begin{aligned} & \int_{-\infty}^{\alpha_X} \xi f_X(\xi) d\xi - \int_{-\infty}^{\alpha_Y} \xi f_Y(\xi) d\xi \\ = & \int_{-\infty}^{\alpha_X} \xi (f_X(\xi) - f_Y(\xi)) d\xi - \int_{\alpha_X}^{\alpha_Y} \xi f_Y(\xi) d\xi \\ = & \xi [F_X(\xi) - F_Y(\xi)] \Big|_{-\infty}^{\alpha_X} - \int_{-\infty}^{\alpha_X} (F_X(\xi) - F_Y(\xi)) d\xi - \int_{\alpha_X}^{\alpha_Y} \xi f_Y(\xi) d\xi \\ = & \alpha_X [F_X(\alpha_X) - F_Y(\alpha_X)] - 0 - \int_{-\infty}^{\alpha_X} (F_X(\xi) - F_Y(\xi)) d\xi - \int_{\alpha_X}^{\alpha_Y} \xi f_Y(\xi) d\xi \\ \leq & \alpha_X (F_X(\alpha_X) - F_Y(\alpha_X)) - \int_{\alpha_X}^{\alpha_Y} \xi f_Y(\xi) d\xi \\ & (\text{according to SSD, the integral of } \int_{-\infty}^{\alpha_X} (F_X(\xi) - F_Y(\xi)) d\xi \\ & \text{must be nonnegative.}) \\ = & \alpha_X (F_X(\alpha_X) - F_Y(\alpha_X)) - \xi F_Y(\xi) \Big|_{\alpha_X}^{\alpha_Y} + \int_{\alpha_X}^{\alpha_Y} F_Y(\xi) d\xi \\ \leq & \beta \alpha_X - \alpha_X F_Y(\alpha_X) - \alpha_Y F_Y(\alpha_Y) + \alpha_X F_Y(\alpha_X) + \beta(\alpha_Y - \alpha_X) \quad (\text{Since } F_Y \leq \beta \text{ on } [\alpha_X, \alpha_Y]) \\ = & 0, \end{aligned}$$

This leads that $\phi_\beta(X) \leq \phi_\beta(Y)$.

Case 2: $\alpha_X \geq \alpha_Y$.

In this case,

$$F_X(\alpha_Y) \leq F_X(\alpha_X) = \beta.$$

Then,

$$\begin{aligned}
& \int_{-\infty}^{\alpha_X} \xi f_X(\xi) d\xi - \int_{-\infty}^{\alpha_Y} \xi f_Y(\xi) d\xi \\
&= \int_{-\infty}^{\alpha_Y} \xi (f_X(\xi) - f_Y(\xi)) d\xi + \int_{\alpha_Y}^{\alpha_X} \xi f_X(\xi) d\xi \\
&= \xi [F_X(\xi) - F_Y(\xi)]|_{-\infty}^{\alpha_Y} - \int_{-\infty}^{\alpha_Y} (F_X(\xi) - F_Y(\xi)) d\xi + \int_{\alpha_Y}^{\alpha_X} \xi f_X(\xi) d\xi \\
&= \alpha_Y [F_X(\alpha_Y) - F_Y(\alpha_Y)] - 0 - \int_{-\infty}^{\alpha_Y} (F_X(\xi) - F_Y(\xi)) d\xi + \int_{\alpha_Y}^{\alpha_X} \xi f_X(\xi) d\xi \\
&\leq \alpha_Y (F_X(\alpha_Y) - F_Y(\alpha_Y)) - 0 + \int_{\alpha_Y}^{\alpha_X} \xi f_X(\xi) d\xi \\
&= \alpha_Y (F_X(\alpha_Y) - F_Y(\alpha_Y)) + \xi F_X(\xi)|_{\alpha_Y}^{\alpha_X} - \int_{\alpha_Y}^{\alpha_X} F_X(\xi) d\xi \\
&\leq -\beta \alpha_Y + \alpha_Y F_X(\alpha_Y) + \beta \alpha_X - \alpha_Y F_X(\alpha_Y) - \beta (\alpha_X - \alpha_Y) \quad (\text{Since } F_X \leq \beta \text{ on } [\alpha_Y, \alpha_X]) \\
&= 0.
\end{aligned}$$

This also implies $\phi_\beta(X) \leq \phi_\beta(Y)$.

Hence, we have proved that

$$X \succeq_2 Y \quad \Rightarrow \quad \phi_X(\beta) \leq \phi_Y(\beta), \forall \beta \in [0, 1],$$

which implies that the mean-risk model $(E_X, CVaR_X)$ is consistent with SSD. \square

3.2.4 Linear Formulation of SSD

For a general probability distribution, the evaluation of the integral in the definition of SSD can lead to considerable computational difficulty. However, if the distribution is discrete this term can be simplified as is shown next [30].

Changing the order of integration in (3.3) and using Fubini's theorem [98], we have

$$\begin{aligned}
F_\omega^2(\eta) &= \int_{-\infty}^{\eta} P(\omega \leq \zeta) d\zeta \\
&= \int_{-\infty}^{\eta} \int_{-\infty}^{\zeta} dp_\omega(\rho) d\zeta \\
&= \int_{-\infty}^{\eta} \int_{\rho}^{\eta} d\zeta dp_\omega(\rho) \\
&= \int_{-\infty}^{\eta} (\eta - \rho) dp_\omega(\rho) \\
&= P(\omega \leq \eta) E[\eta - \omega | \omega \leq \eta] \\
&= E[(\eta - \omega)_+]
\end{aligned} \tag{3.7}$$

Hence, the SSD defined in (3.4) can be written as

$$E[(\eta - Y)_+] \leq E[(\eta - L)_+], \quad \forall \eta \in \mathbb{R}. \tag{3.8}$$

Notice that this reformulation of SSD has similarity with CVaR as in (1.5). While a CVaR value is only concerned with a single real number η (i.e. the corresponding VaR), a SSD constraint as defined in (3.8) works $\forall \eta \in \mathbb{R}$. This also facilitates the use of SSD instead of CVaR.

To make the problem easier for modelling and computation, consider a relaxed formulation of this constraint valid in the interval $[a, b]$:

$$E[(\eta - Y)_+] \leq E[(\eta - L)_+], \quad \eta \in [a, b]. \quad (3.9)$$

Let $v \in \mathbb{R}$ denote the shortfall. We can show that (3.9) is equivalent to:

$$\begin{cases} Y + v \geq \eta, \quad \eta \in [a, b] \\ E[v] \leq E[(\eta - L)_+], \quad \eta \in [a, b] \\ v \geq 0. \end{cases} \quad (3.10)$$

If L has a discrete probability distribution with realizations l_k , for $k = 1, \dots, K$ and $a \leq l_k \leq b$, then (3.9) can be rewritten as

$$E[(l_k - Y)_+] \leq E[(l_k - L)_+], \quad \forall k = 1, \dots, K. \quad (3.11)$$

Furthermore, if Y has a discrete distribution with realizations y_m , for $m = 1, \dots, M$ and $a \leq y_m \leq b$, with π_m denoting the probability of y_m occurring, (3.10) becomes

$$\begin{cases} y_m + v_{m,k} \geq l_k, \\ \sum_m \pi_m v_{m,k} \leq \hat{l}_k, \\ v_{m,k} \geq 0, \end{cases} \quad (3.12)$$

where $\hat{l}_k = E[(l_k - L)_+]$. It is easy to see that (3.12) are linear.

3.2.5 Interval Second-order Stochastic Dominance and Chance Constraints

The interval SSD was first introduced by Noyan et al. in [96] and proved to be a sufficient as well as a necessary condition of FSD. Here, we will consider a relaxed interval SSD in the discrete case as an intermediate stochastic dominance relation between FSD and SSD, i.e. a weaker condition than FSD, but stronger than SSD. This relaxed interval SSD can be developed to construct chance constraints for underfunding control in ALM.

We say that a random variable Y dominates another L by interval second-order stochastic dominance (ISSD) if and only if:

$$E[(\eta_2 - Y)_+] - E[(\eta_1 - Y)_+] \leq E[(\eta_2 - L)_+] - E[(\eta_1 - L)_+], \quad (3.13)$$

for any $\eta_1, \eta_2 \in \mathbb{R}$ and $\eta_1 \leq \eta_2$.

The proposition below establishes a relation between FSD and ISSD. It was first proved in [96] in the case of discrete probability distributions. We shall prove it in a general form.

Proposition 3. *$Y \succeq_1 L$ if and only if Y dominates L by ISSD.*

Proof. The proof of necessity is simple. If $Y \succeq_1 L$, then for any given $\eta_1 \leq \eta_2$ and t , $\eta_1 \leq t \leq \eta_2$,

$$0 \leq F_Y^1(t) \leq F_L^1(t).$$

Hence, integrating

$$\int_{\eta_1}^{\eta_2} F_Y^1(t) dt \leq \int_{\eta_1}^{\eta_2} F_L^1(t) dt. \quad (3.14)$$

Following Equation (3.7), we observe that Equation (3.14) is equivalent to the definition of ISSD, i.e. Equation (3.13).

We prove the sufficiency by contradiction. Suppose that there exists t^* such that

$$F_Y^1(t^*) > F_L^1(t^*).$$

Let $[a^*, b^*]$ be an interval such that $t^* \in [a^*, b^*]$ and

$$\begin{aligned} a^* &= \inf\{a : F_Y(t) > F_L(t), t \in [a, t^*]\} \\ b^* &= \sup\{b : F_Y(t) > F_L(t), t \in [t^*, b]\}. \end{aligned}$$

Since both distribution functions F_Y and F_L are semi-continuous, it follows that $a^* < b^*$. Then, we have

$$\int_{a^*}^{b^*} F_Y^1(\alpha) d\alpha > \int_{a^*}^{b^*} F_L^1(\alpha) d\alpha$$

which violates the definition of ISSD. The sufficiency is proved. \square

If Y and L both have discrete probability distributions with realisations $y_1 < y_2 < \dots < y_M$, and $l_1 < l_2 < \dots < l_K$, the ISSD condition can be written as:

$$E[(l_k - Y)_+] - E[(y_m - Y)_+] \leq E[(l_k - L)_+] - E[(y_m - L)_+], \quad (3.15)$$

for all $m \in \{1, \dots, M\}$ and $k \in \{1, \dots, K\}$ such that $l_k \geq y_m$ and

$$\{l_1, \dots, l_K, y_1, \dots, y_M\} \cap (y_m, l_k) = \emptyset,$$

where (y_m, l_k) is the open interval with endpoints y_m and l_k [96].

Incorporating constraints (3.15) into the ALM model (3.1) leads to a mixed integer formulation, where boolean variables are induced by the conditional expectation in ISSD equations. Hence, we consider a relaxed form of ISSD in the case of discrete

probability distributions:

$$E[(l_k - Y)_+] - E[(l_{k-1} - Y)_+] \leq E[(l_k - L)_+] - E[(l_{k-1} - L)_+], \quad k \in 2, \dots, K, \quad (3.16)$$

where l_k , $k = 2, \dots, K$, are the realisations of L and l_0 is any real number such that $l_0 < l_1$, and denote the above relation of Y and L as

$$Y \succeq_{1\frac{1}{2}} L.$$

It is easy to prove that the relaxed ISSD is weaker than FSD but stronger than SSD, i.e.

$$FSD \Rightarrow \text{Relaxed ISSD} \Rightarrow SSD. \quad (3.17)$$

The first implication was proved in [96]. We give a full picture of these three dominance relations in the following proposition.

Proposition 4. *If Y dominates L by FSD, then $Y \succeq_{1\frac{1}{2}} L$; If $Y \succeq_{1\frac{1}{2}} L$, then Y dominates L by SSD.*

Proof. By Proposition 3, if FSD is true, ISSD is satisfied, which is sufficient for relaxed ISSD.

If relaxed ISSD is satisfied, we have

$$\int_{l_{k-1}}^{l_k} F_Y^1(t) dt \leq \int_{l_{k-1}}^{l_k} F_L^1(t) dt,$$

for $k = 1, \dots, K$. Since $F_L^1(t) = 0$, for any t such that $l_0 < t < l_1$,

$$\int_{l_0}^{l_1} F_Y^1(t) dt \leq \int_{l_0}^{l_1} F_L^1(t) dt = 0$$

from which we deduce that $F_Y(t) = 0$, a.e., for $t < l_1$. Hence, for any real number $\eta \leq l_1$,

$$\int_{-\infty}^{\eta} F_Y^1(t) dt \leq \int_{-\infty}^{\eta} F_L^1(t) dt = 0.$$

Also, for $k = 1, \dots, K$,

$$\begin{aligned} \int_{-\infty}^{l_k} F_Y^1(t) dt &= \int_{-\infty}^{l_1} F_Y^1(t) dt + \sum_{j=1, \dots, k-1} \int_{l_j}^{l_{j+1}} F_Y^1(t) dt \\ &\leq 0 + \sum_{j=1, \dots, k-1} \int_{l_j}^{l_{j+1}} F_L^1(t) dt \\ &= \int_{-\infty}^{l_k} F_L^1(t) dt. \end{aligned}$$

Now suppose that there exists $\eta \in [l_k, l_{k+1}]$ such that

$$\int_{-\infty}^{\eta} F_Y^1(t) dt > \int_{-\infty}^{\eta} F_L^1(t) dt.$$

Since

$$\int_{-\infty}^{l_k} F_Y^1(t) dt \leq \int_{-\infty}^{l_k} F_L^1(t) dt,$$

we have

$$\int_{l_k}^{\eta} F_Y^1(t) dt > \int_{l_k}^{\eta} F_L^1(t) dt. \quad (3.18)$$

In addition, for $t \in [l_k, l_{k+1})$, $F_L^1(t) = F_L^1(l_k)$. From (3.18), using monotonicity of F_Y^1 ,

$$F_Y^1(\eta) > F_L^1(l_k).$$

As a result,

$$\int_{\eta}^{l_{k+1}} F_Y^1(t) dt > \int_{\eta}^{l_{k+1}} F_L^1(t) dt. \quad (3.19)$$

(3.18) and (3.19) together imply

$$\int_{l_k}^{l_{k+1}} F_Y^1(t) dt > \int_{l_k}^{l_{k+1}} F_L^1(t) dt,$$

which contradicts the relaxed ISSD condition. Therefore, for all $\eta \in [l_1, l_K]$, SSD is satisfied.

For $\eta > l_K$,

$$\begin{aligned} \int_{-\infty}^{\eta} F_L^1(t) dt &= \int_{-\infty}^{l_K} F_L^1(t) dt + \int_{l_K}^{\eta} F_L^1(t) dt \\ &= \int_{-\infty}^{l_K} F_L^1(t) dt + \int_{l_K}^{\eta} 1 dt \\ &\geq \int_{-\infty}^{l_K} F_Y^1(t) dt + \int_{l_K}^{\eta} F_Y^1(t) dt. \end{aligned}$$

Hence, Y dominates L by SSD. \square

An interesting question arises whether any reverse implication to (3.17) holds. Two examples are given below to illustrate that the other directions of the relations are not true. The first demonstrates that the relaxed ISSD does not imply FSD and the second shows that SSD does not imply the relaxed ISSD.

Example 1. Consider two assets L and Y with the following probability distributions of returns: $P(L = 100) = \frac{1}{3}$, $P(L = 200) = \frac{1}{3}$, $P(L = 300) = \frac{1}{3}$; $P(Y = 150) = \frac{1}{2}$,

$P(Y = 300) = \frac{1}{2}$. For these distributions, we find:

$$E[(\eta - L)_+] = \begin{cases} 0, & \eta \leq 100 \\ \frac{1}{3}(\eta - 100), & 100 < \eta \leq 200 \\ \frac{2}{3}(\eta - 200) + \frac{1}{3}(200 - 100), & 200 < \eta \leq 300 \\ (\eta - 300) + \frac{2}{3}(300 - 200) + \frac{1}{3}(200 - 100), & 300 < \eta \end{cases}$$

$$E[(\eta - Y)_+] = \begin{cases} 0, & \eta \leq 150 \\ \frac{1}{2}(\eta - 150), & 150 < \eta \leq 300 \\ (\eta - 300) + \frac{1}{2}(300 - 150), & 300 < \eta \end{cases}$$

and collect the values of $E[(l_k - X)_+] - E[(l_{k-1} - X)_+]$ for both variables L and Y for all intervals $(l_{k-1}, l_k]$ in the table below:

$E[l_k - X]_+ - E[l_{k-1} - X]_+$	$[0, 100]$	$(100, 200]$	$(200, 300]$
X=L	0	33.3	66.7
X=Y	0	25	50

Table 3.1: The relaxed ISSD values of assets L and Y .

Obviously, inequality (3.16) is always satisfied hence the relaxed ISSD is satisfied, i.e. $Y \succeq_{1\frac{1}{2}} L$. However, $P(L \leq 150) < P(Y \leq 150)$, which means FSD is violated. \square

Example 2. Consider two assets L and Y , where L is the same as in Example 1. Asset Y has three possible returns: $P(Y = 150) = \frac{1}{2}$, $P(Y = 200) = \frac{1}{4}$ and $P(Y = 300) = \frac{1}{4}$. Y dominates L by SSD but Y does not dominate L by relaxed ISSD, because

$$F_\omega^2 = E[(\eta - \omega)_+] = \int_{-\infty}^{\eta} F_\omega(\xi) d\xi,$$

$$F_L^2(\eta) = E[(\eta - L)_+] = \begin{cases} 0, & \eta \leq 100 \\ \frac{1}{3}(\eta - 100), & 100 < \eta \leq 200 \\ \frac{2}{3}(\eta - 200) + \frac{1}{3}(200 - 100), & 200 < \eta \leq 300 \\ (\eta - 300) + \frac{2}{3}(300 - 200) + \frac{1}{3}(200 - 100), & 300 < \eta \end{cases}$$

$$F_Y^2(\eta) = E[(\eta - Y)_+] = \begin{cases} 0, & \eta \leq 150 \\ \frac{1}{2}(\eta - 150), & 150 < \eta \leq 200 \\ \frac{3}{4}(\eta - 200) + \frac{1}{2}(200 - 150), & 200 < \eta \leq 300 \\ (\eta - 300) + \frac{3}{4}(300 - 200) + \frac{1}{2}(200 - 150), & 300 < \eta \end{cases}$$

illustrating that $E[(\eta - L)_+] \geq E[(\eta - Y)_+]$, while

$$E[(300 - L)_+] - E[(200 - L)_+] = \frac{200}{3} \leq E[(300 - Y)_+] - E[(200 - Y)_+] = 75.$$

\square

Below we prove one more technical result regarding relaxed ISSD which has important consequences for a practical way of modelling relaxed ISSD constraints as explained in the rest of this section.

Proposition 5. *Let Y and L be random variables, whose probability distributions are discrete with realisations y_1, \dots, y_M and l_1, \dots, l_K , respectively. Let Y dominate L by relaxed ISSD. If there exists $k \in \{1, \dots, K-1\}$, such that*

$$\{y_1, \dots, y_M\} \cap (l_k, l_{k+1}) = \emptyset,$$

then $F_Y^1(t) \leq F_L^1(t)$ for all $t \in [l_k, l_{k+1}]$

Proof. For any k such that

$$\{y_1, \dots, y_M\} \cap (l_k, l_{k+1}) = \emptyset,$$

$F_Y^1(t) = F_Y^1(l_k)$, $t \in [l_k, l_{k+1})$. Then by the relaxed ISSD relation,

$$\begin{aligned} \int_{l_k}^{l_{k+1}} F_Y^1(t) dt &= F_Y^1(l_k)(l_{k+1} - l_k) \leq \int_{l_k}^{l_{k+1}} F_L^1(t) dt = F_L^1(l_k)(l_{k+1} - l_k) \\ &\Rightarrow F_Y^1(l_k) \leq F_L^1(l_k). \end{aligned}$$

□

Remark 3.1. *By comparing relaxed ISSD and ISSD which is equivalent to FSD, we can see the relaxation is at the points of y_m . Assume relaxed ISSD is true. From the above proposition, the FSD is satisfied in any interval $[l_k, l_{k+1})$ which does not contain any y_m . Actually, even if y_m appears in this interval, FSD still holds if $F_Y^1(y_m) \leq F_L^1(l_k)$. FSD is violated only in the interval in which the probability of Y jumps over the probability of the benchmark L . And this violation will not transfer to the next interval because of relaxed ISSD.*

Proposition 5 opens a way to express chance constraints in LP form by imposing relaxed ISSD constraints. Assume L is a benchmark with discrete distribution and the portfolio Y dominates L by relaxed ISSD, and let $l_k < l_{k+1}$ be two neighbouring realisations of the benchmark. If $[l_k, l_{k+1}]$ is such that the portfolio will not have any realization in this interval, then

$$P(Y \leq t) \leq P(L \leq t), \forall t \in [l_k, l_{k+1}]$$

Hence, the probability of the portfolio can be constrained for those values in such intervals. There is an issue of how to guarantee the existence of such intervals. We address this problem below.

The risk control in ALM modelling reflects concerns about the underfunding which is the amount of unsatisfied liability. Bogentoft, et al. [14] applied CVaR to control

the return of the pension fund with certain percentage to cover the liability. While it is difficult and costly to avoid any underfunding at all, it seems highly desirable to limit the probability that any underfunding happens. We will show how to express such probability constraints in LP form. Suppose the portfolio is expected to satisfy the following chance constraint:

$$P(\text{final wealth} - \text{liability} < 0) \leq \alpha, \quad (3.20)$$

where α is a given threshold. We can construct such an interval $[\theta_1, \theta_2]$ that the following two equations

$$\text{final wealth} - \text{liability} < \theta_1 \quad (3.21)$$

$$\text{final wealth} - \text{liability} < \theta_2 \quad (3.22)$$

are equivalent to

$$\text{final wealth} - \text{liability} < 0. \quad (3.23)$$

For example, if it is the same to the fund manager in practice to have either no underfunding or an underfunding of £1, then this interval can be $[-1, 0]$. We assume that such an interval always exists. Suppose the return of the portfolio is modelled by M scenarios. A benchmark L can be constructed satisfying the following conditions:

- The benchmark value has K realizations and $K > M + 1$;
- Among the K realizations, at least $M + 1$ are allocated in the interval $[\theta_1, \theta_2]$, with θ_1 and θ_2 defined as above; and
- Last but most important, $P(L - \text{Liability} < 0) \leq \alpha$.

If a portfolio outperforms such a benchmark by relaxed ISSD, there must be an interval $[l_k, l_{k+1}) \subset [\theta_1, \theta_2]$, where the portfolio value has no realization. Then by Proposition 5, this portfolio has return below l_{k+1} with probability less than α . While there is no difference to the fund manager to have an underfunding of l_{k+1} or 0, the chance constraint of the underfunding is successfully satisfied. For multiple chance constraints, separate relaxed ISSD constraints can be applied and the derivation is the same as in the single case.

3.3 Multi-Stage ALM Model with SSD and Relaxed ISSD Constraints

Now, we will apply SSD and relaxed ISSD in the multi-stage ALM model to control risk. Either SSD or relaxed ISSD can be incorporated in the model independently. Both SSD and relaxed ISSD constraints are set at each stage: overall portfolio returns

are required to dominate a benchmark by SSD; relaxed ISSD constraint guarantees that the portfolio value minus liabilities dominates a benchmark by relaxed ISSD.

In addition to the notation listed in Section 3.1, new notation is introduced to construct the stochastic dominance constraints in the model as follows:

New notation

τ_t^l : the values of benchmark performance at stage t used for the SSD constraint, $l = 1, \dots, L$;

μ_t^k : the values of benchmark performance at stage t used for the relaxed ISSD constraint, $k = 1, \dots, K$. If we want the probability of underfunding to be less than or equal to α , then these values are set such that the probability of this benchmark being negative is equal to α ;

$\hat{\tau}_t^l = E[(\tau_t^l - \tau_t)_+]$, $l = 1, \dots, L$;

$\hat{\mu}_t^k = E[(\mu_t^k - \mu_t)_+]$, $k = 1, \dots, K$;

$\mathbf{z}_{j,t}^l$: shortfall of the portfolio in node j at stage t compared to l th value of the benchmark at stage t in the SSD constraint;

$\mathbf{v}_{j,t}^k$: shortfall of the portfolio in node j at stage t compared to k th value of the benchmark at stage t in the relaxed ISSD constraint.

Model (3.24) is obtained by including SSD and ISSD constraints into the model presented in (3.1), where all scenarios together at each stage are restricted to one SD constraint. Equations (3.24f) and (3.24g) are SSD constraints that restrict the return of the portfolio so that it dominates the benchmark τ by SSD; while (3.24h), (3.24i) and (3.24j) are relaxed ISSD constraints which guarantee that the value of the portfolio minus the amount of the liability dominates the benchmark μ by relaxed ISSD and so control the probability of underfunding:

$$\max \sum_{j_T \in N_T} \pi_{j_T}^T \left(\sum_{i \in I} (1 - \gamma) W_i x h_{i,j_T}^T + c_{j_T}^T - \lambda b_{j_T}^T \right) \quad (3.24a)$$

$$s.t. \quad (1 + \gamma) \sum_{i \in I} W_i x h_{i,0}^0 + c_0 = G - A_0 + D_0 \quad (3.24b)$$

$$(1 + \gamma) \sum_{i \in I} W_i x b_{i,j_t}^t + c_{j_t}^t = (1 - \gamma) \sum_{i \in I} W_i x s_{i,j_t}^t + (1 + R_{c,j_t}^t) c_{a(t,j_t)}^{t-1} - A_{j_t}^t + D_{j_t}^t, \quad (3.24c)$$

$$j_t = 1, \dots, n_t, \quad t = 1, \dots, T,$$

$$(1 + R_{i,j_t}^t) x h_{i,a(t,j_t)}^{t-1} + x b_{i,j_t}^t - x s_{i,j_t}^t = x h_{i,j_t}^t, \quad (3.24d)$$

$$i \in I, \quad j_t = 1, \dots, n_t, \quad t = 1, \dots, T,$$

$$\sum_{i \in I} (1 - \gamma) W_i x h_{i,j_T}^T + c_{j_T}^T + b_{j_T}^T \geq \psi A_{j_T}^T, \quad j_T = 1, \dots, n_T \quad (3.24e)$$

$$\sum_{i \in I} (1 + R_{i,j_t}^t) W_i x h_{i,a(t,j_t)}^{t-1} + (1 + R_{c,j_t}^t) c_{a(t,j_t)}^{t-1} + z_{j_t,t}^l \geq \tau_t^l, \quad (3.24f)$$

$$j_t = 1, \dots, n_t, \quad t = 1, \dots, T, \quad l = 1, \dots, L,$$

$$\sum_{j_t \in N_t} \pi_{j_t}^t z_{j_t,t}^l \leq \hat{\tau}_t^l, \quad t = 1, \dots, T, \quad l = 1, \dots, L, \quad (3.24g)$$

$$\sum_{i \in I} (1 + R_{i,j_t}^t) W_i x h_{i,a(t,j_t)}^{t-1} + (1 + R_{c,j_t}^t) c_{a(t,j_t)}^{t-1} - \psi_{A_{j_t,t}} + v_{j_t,t}^k \geq \mu_t^k, \quad (3.24h)$$

$$j_t = 1, \dots, n_t, \quad t = 1, \dots, T, \quad k = 1, \dots, K,$$

$$\sum_{j_t \in N_t} \pi_{j_t}^t v_{j_t,t}^k - \sum_{j_t \in N_t} \pi_{j_t}^t v_{j_t,t}^{k-1} \leq \hat{\mu}_t^k - \hat{\mu}_t^{k-1}, \quad t = 1, \dots, T, \quad k = 1, \dots, K, \quad (3.24i)$$

$$\sum_{j_t \in N_t} \pi_{j_t}^t v_{j_t,t}^1 \leq \hat{\mu}_t^1, \quad t = 1, \dots, T, \quad (3.24j)$$

$$x h_{i,j_t}^t \geq 0, \quad x s_{i,j_t}^t \geq 0, \quad x b_{j_t}^t \geq 0, \quad b_{j_t}^T \geq 0, \quad z_{j_t,t}^l \geq 0, \quad v_{j_t,t}^k \geq 0$$

$$x h_{j_t}^t, \quad x s_{j_t}^t, \quad x b_{j_t}^t, \quad z_{j_t,t}^l, \quad v_{j_t,t}^l \in \mathbb{R}^m$$

$$i \in I = \{1, \dots, m\}, \quad j \in N_t = \{1, \dots, n_t\}, \quad t = 1, \dots, T,$$

$$l = 1, \dots, L, \quad k = 1, \dots, K.$$

The objective (3.24a) is to maximise the final wealth considering the penalty of underfunding. In addition to the initial budget constraint (3.24b), asset balance constraint (3.24c), inventory constraint (3.24d) and underfunding definition in (3.24e), risk control is modelled by (3.24f)–(3.24j). (3.24f) and (3.24g) require the portfolio dominate benchmark τ by SSD. In (3.24h)–(3.24j) the total value of the portfolio deducted by the liability dominates benchmark μ by relaxed ISSD, where μ is constructed according to chance requirement of underfunding. They are all in discrete case and thus the equations are linear. Now we can see that, by incorporating SSD in the ALM model, the risk incurring overall performance is controlled by requesting that our portfolio outperforms the benchmark by SSD; by incorporating relaxed ISSD, the risk of underfunding is controlled in terms of chance constraints.

3.4 Numerical Results

The models discussed in this work are applicable in practice. We first demonstrate the advantages of taking stochastic dominance constraints into account using a small example, followed with an out-of-sample backtest. Then we will show how real-world problems can be solved. We use the structure-exploiting interior point solver OOPS [57] to solve these problems and compare its performance with that of the general-purpose commercial optimizer CPLEX 10.0 on a number of medium scale test examples.

3.4.1 A Model Example

Consider a small investment problem with 2 stages and 4 stocks (named A, B, C, D) to be chosen from. One stage corresponds to one day. There are 4 branches at the first stage and 2 branches from each node of the second stage. Both asset returns and

liabilities are random. The asset returns are from FTSE100 and the probabilities are made up. The returns in per cent of the 4 stocks and liabilities in monetary value are shown in Table 3.2 and Table 3.3 with the probabilities in brackets and the other parameters are presented in Table 3.4:

	Stocks					Stocks			
1st Stg	A	B	C	D	2nd Stg	A	B	C	D
1 (0.5)	0.0145	-0.1020	-0.0305	0.2299	1 (0.40)	0.1145	-0.2020	-0.0305	0.0299
					2 (0.10)	-0.1060	0.2450	0.0341	0.0167
2 (0.2)	0.0056	0.2050	0.1041	-0.0236	3 (0.16)	0.1145	-0.2020	-0.0305	0.0299
					4 (0.04)	-0.1060	0.2450	0.0341	0.0167
3 (0.2)	-0.0113	0.0007	-0.0287	0.1658	5 (0.16)	0.1145	-0.2020	-0.0305	0.0299
					6 (0.04)	-0.1060	0.2450	0.0341	0.0167
4 (0.1)	0.1573	-0.0286	0.0645	-0.0742	7 (0.08)	0.1145	-0.2020	-0.0305	0.0299
					8 (0.02)	-0.1060	0.2450	0.0341	0.0167

Table 3.2: Rate of Return of the assets in per cent.

1st Stg	1 (0.5)		2 (0.2)		3 (0.2)		4 (0.1)	
	10200		10000		10200		10000	
2nd Stg	1 (0.4)	2 (0.1)	3 (0.16)	4 (0.04)	5 (0.16)	6 (0.04)	7 (0.08)	8 (0.02)
	11220	10000	11000	10000	11220	10000	11000	10000

Table 3.3: Liabilities in value.

Description	Parameter	Value
number of assets	m	4
number of leaf nodes	n_T	8
number of SSD benchmarks	K_1	1
number of relaxed ISSD benchmarks	K_2	1
length of investment horizon	T	2
initial budget	A_0	10000
penalty coefficient for underfunding at horizon	λ	2
lower bound of funding ratio	ψ	1.01
transaction fee ratio	γ	0.03

Table 3.4: Typical parameter values.

We obtain the optimal investment strategy using 3 models. In the first one, (i), the underfunding is penalized in the objective without any SD constraint, presented as Equations (3.24a)–(3.24e). In the second one, (ii), an SSD constraint is added to the first model (i) to restrict the portfolio to outperform a benchmark at the first stage, presented as Equations (3.24a)–(3.24g). As the third model, (iii), we apply the full model (3.24) for this problem presented as Equations (3.24a)–(3.24j), where the probability of underfunding at the final (second) stage is restricted to be less than 5%

Number	1	2	3
Value	-10	0	10
Probabilities	0.2	0.6	0.2

(a) 1st stage benchmark

Number	1	2	3	4	5	6	7	8	9
Value	-10	-8	-6	-4	-2	-1	-0.9	-0.8	-0.7
Probabilities	0.01	0.01	0.005	0.005	0.005	0.005	0.001	0.001	0.001
Number	10	11	12	13	14	15	16	17	–
Value	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0	10	–
Probabilities	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.95	–

(b) 2nd stage benchmark

Table 3.5: Benchmark values and probability distributions.

by relaxed ISSD constraints, with all other features the same as for the second model. The benchmarks used for SSD and relaxed ISSD constraints are listed in Table 3.5. As we can see the probability of the 2nd stage benchmark less or equal to 0 is 5%.

Results are summarised in Table 3.6, where the models are solved by AMPL&CPLEX. Model (i) suggested investing only in assets *A* and *D*, while both models (ii) and (iii) included also asset *B* with slight differences in the weights of each asset respectively. Assets *A* and *D* have better performance in terms of expected return compared with the other two. However, the inclusion of asset *B* can lead to better diversification. From the results presented in Table 3.6, we can see that taking SSD constraints into account can half the probability of underfunding while the expected return is reduced by 45%. Relaxed ISSD together with SSD can effectively reduce the probability of underfunding to merely 2% while the return is the same.

Models		Asset Allocation				Expected Portfolio Return	Prob(underfunding)
		A	B	C	D		
(i)	No SD	26.6%	0	0	73.4%	10.18%	22%
(ii)	SSD	39.3%	18.2%	0	42.5%	6.25%	6%
(iii)	SSD + rISSD	41.5%	15.6%	0	42.9%	5.70%	2%

Table 3.6: Portfolio properties generated from 3 models: portfolio composition, expected return and the probability of underfunding.

Figure 3.2 and Figure 3.3 illustrate the second-order stochastic dominance relations of the portfolios generated by the 3 models compared to the benchmark at the first stage and the second stage, respectively. The performance compared in the figures are the value of the portfolio minus the liabilities. Line above is dominated by line entirely below. We can see from Figure 3.2 that both portfolios by model (ii) and model (iii) dominate the benchmark by SSD, while the model (i) portfolio does not. In Figure 3.3,

we specially show the relations in the interval $[-10, 10]$, because such relation is not clear in the graph with larger x-axis range. For the same reason, model (i) portfolio which does not dominate the benchmark is not shown in this figure. At the second stage, only model (iii) portfolio dominates the benchmark.

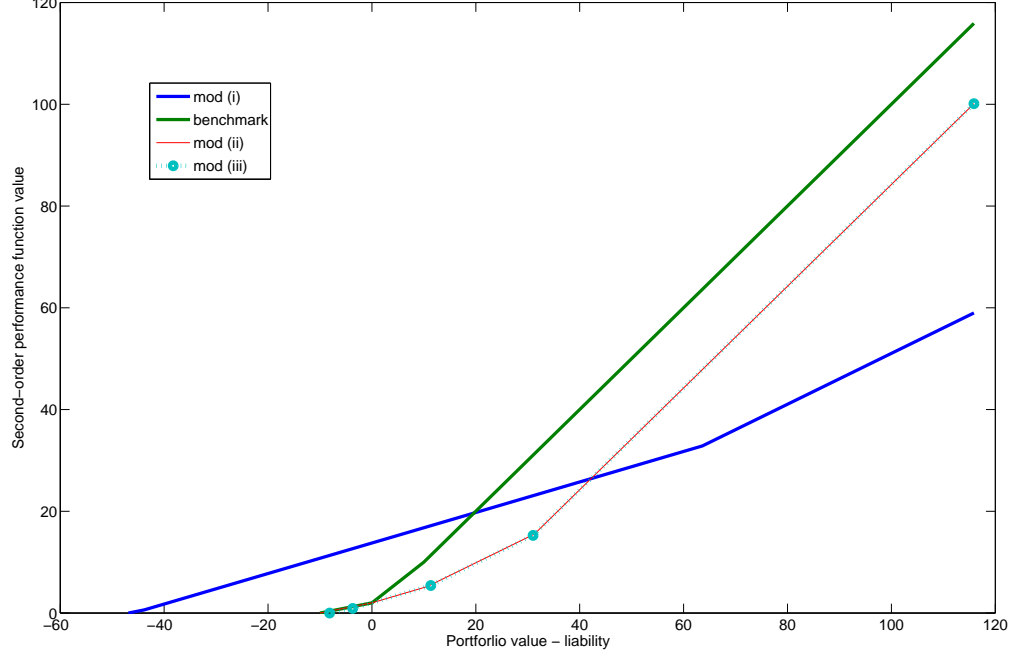


Figure 3.2: Stochastic dominance relations of portfolio performances at the first stage.

3.4.2 Backtesting

Assume a fund management project with the initial wealth of £100,000 which can be invested in the stocks in FTSE100 and as deposits in money market with certain interest. Benchmarks are constructed as the worst portfolio that can be accepted. All the parameters in the backtest are the same as in Table 3.4 except the number of assets and the number of nodes, which are 102 and 80 respectively. We used daily rate of return as the first stage scenarios and monthly rate of return in the same time period as the second stage scenarios. The backtest is run with 80 rolling time windows. The model first learns from 80-day market data to generate the optimal portfolio, then we apply this portfolio strategy to the next 80 days and compare its performance with a passive investment strategy only on FTSE100 Index. Then roll this 160-day time window by one day to the next and repeat above computations. For example in the first window, we run the model with market data from 09/10/2007 to 16/02/2008 (80 days), to generate an optimal portfolio strategy; then see how this portfolio performs from 19/02/2008 to 11/06/2008 (the following 80 days, 17/02/2008 and 18/02/2008 are

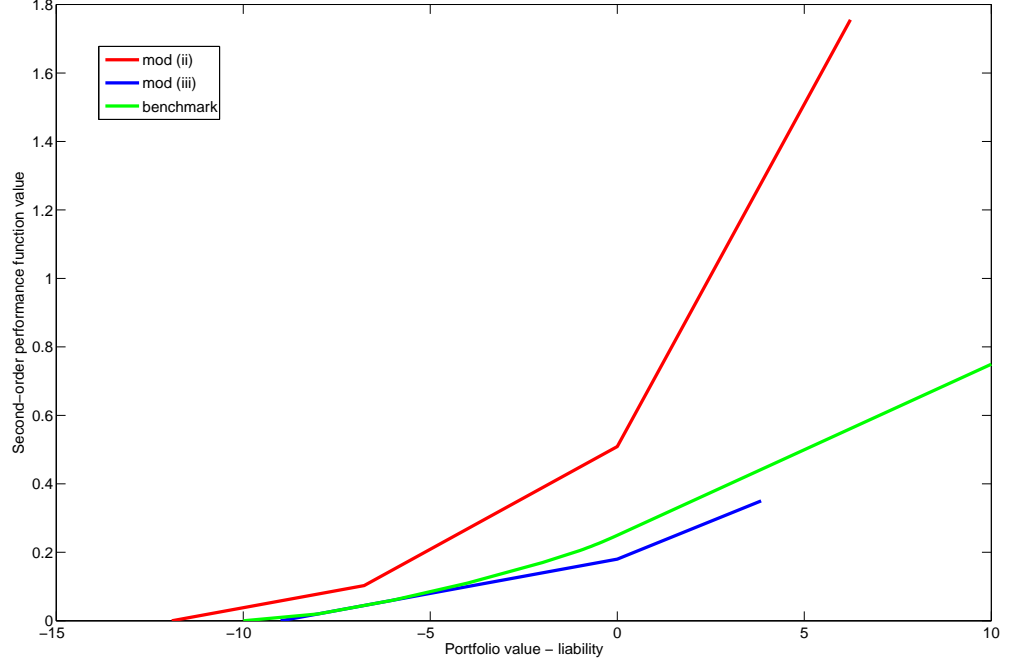


Figure 3.3: Stochastic dominance relations of portfolio performances at the second stage.

weekends). In Fig 3.4, the results are shown with a thick solid line for the portfolio and a dotted line for FTSE100 Index, where the curve of the portfolio is generally above the Index.

For each 80-day time window, we count percentage of days that underfunding occurs and use them as an indicator of possibility of underfunding. Fig 3.5 shows, for the 80 time windows, the percentages of days out of 80 that underfunding over 5% occurs, and the percentages of days out of 80 that underfunding over 10% occurs are in Fig 3.6, for both the portfolio and the FTSE100 Index. In Fig 3.5, we can see that, until 04/04/2008 the underfunding of the portfolio over 5% appeared less frequently than that of the Index, which is below 10% of 80 days. Then the curve of the portfolio jumped above the Index. This jump was due to the big recession of the market starting from 19/05/2008. Fig 3.7 shows the performance of the portfolio, the Index and the Markowitz strategy over those 80 days from 04/04/2008. The portfolio performs relatively steadily compared to the Index, i.e. the underfunding is below 10%, although the underfunding exists through the whole period. The worst performance of the portfolio in both Fig 3.5 and 3.6 appears around 20/05/2008, similarly to the Index, when the market was at a turning point and started decrease along the way until touching the 21-month-low on 15/07/2008.

Through the whole test, the portfolio generated by the model presents a relative steady performance compared to FTSE100 Index. For example on 28/04/2008, two



Figure 3.4: Value of the portfolio and FTSE100 Index in first time window (19/02/2008-11/06/2008).

lines in Fig 3.5 cross, which means the percentages of days with underfunding over 5% are the same for both the portfolio and the Index investment. However, the possibility of the portfolio underfunding over 10% corresponding to that day is 5% as shown in Fig 3.6, significantly smaller than 37.5% of the Index. That 80-day performance is illustrated in Fig 3.8. Among the 80 time windows, there are 58 windows (72.5%) when the percentage of days with portfolio underfunding over 5% is smaller than that of the Index as illustrated in Fig 3.5, and 75 windows (93.75%) for underfunding over 10% as in Fig 3.6.

3.4.3 Numerical Efficiency

The ALM stochastic programming model (3.24) proposed in the previous section has the constraint matrix structure shown in Figure 3.9. Each diagonal block composed of small A and B matrices corresponds to an initial branch in the event tree. It contains the inventory, cash balance and underfunding definition at the last stage. The most right column contains the coefficients of the first stage variables and the bottom diagonal block contains the initial budget constraint. The bottom border corresponds to the stochastic dominance constraints linking all the nodes of a given stage together. For multi-stage problems, the small diagonal A -blocks are themselves structured. This nested bordered block-diagonal structure can be efficiently exploited by OOPS [55,

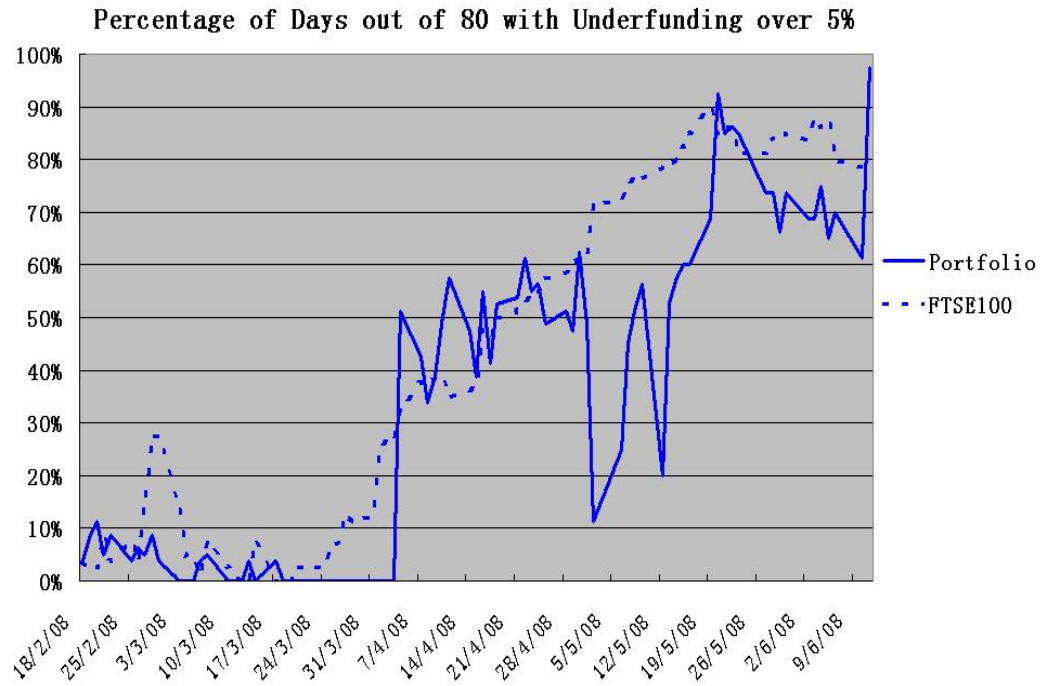


Figure 3.5: Percentage of days with underfunding over 5%.

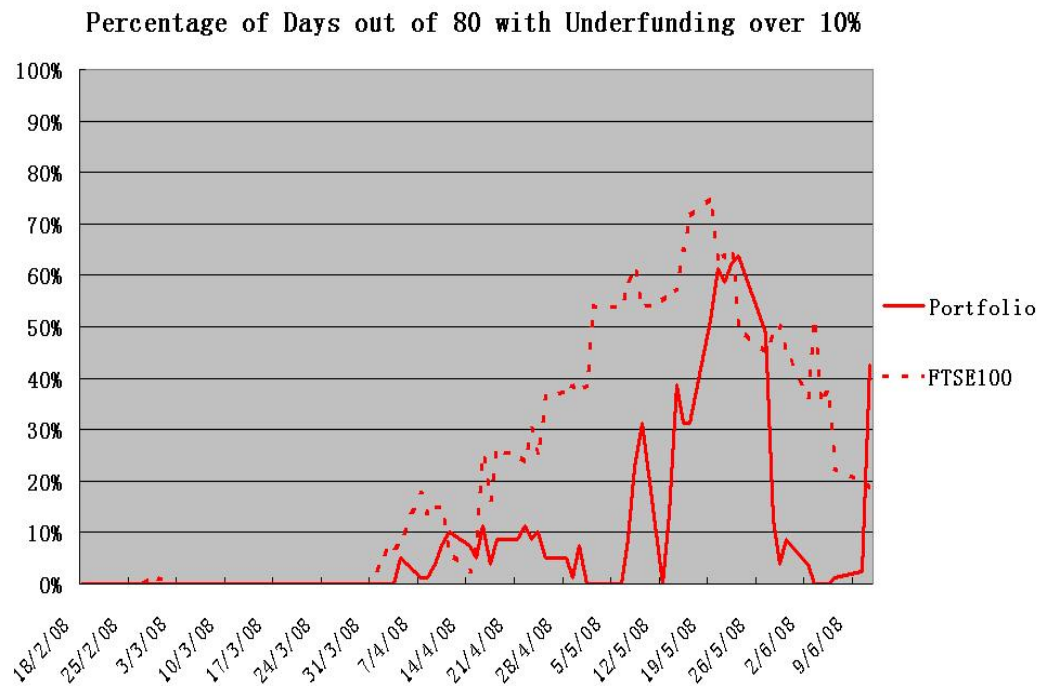


Figure 3.6: Percentage of days with underfunding over 10%.

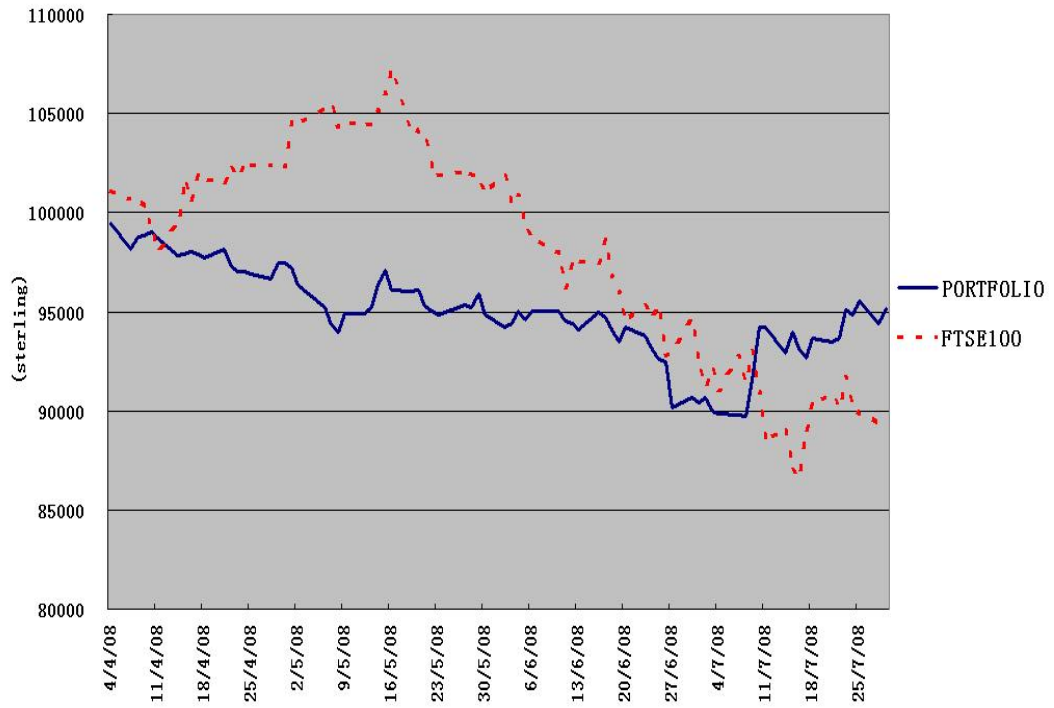


Figure 3.7: Value of the portfolio and FTSE100 Index over 04/04/2008-29/07/2008.

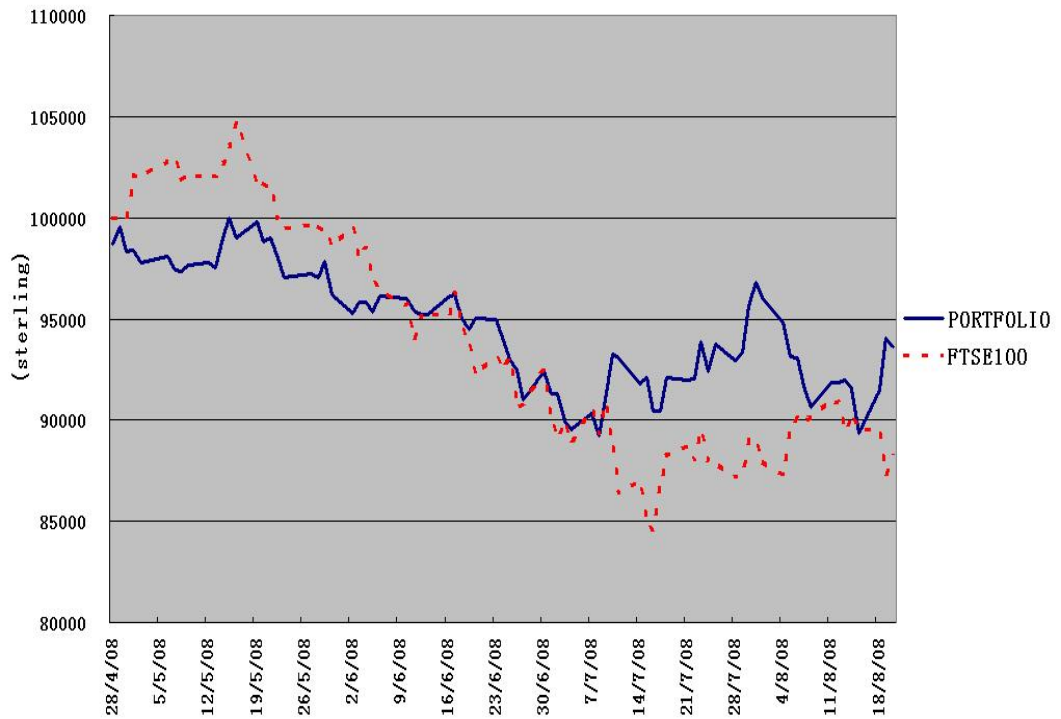


Figure 3.8: Value of the portfolio and FTSE100 Index over 28/04/2008-20/08/2008.

Problem	Stages	Blocks	Assets	Bnmk	Total Nodes	Constraints	Variables
	T	B	I	L	$ N = \sum_{t=0}^{T-1} N_t$	$(I+L+2) N $	$(3I+L+2) N $
ALM1a	2	80	64	20	81	6966	17334
ALM1b	2	40	128	20	41	6150	16646
ALM1c	2	80	128	20	81	12150	32886
ALM1d	2	160	128	20	161	24150	65366
ALM2a	2	80	64	40	81	8586	18954
ALM2b	2	40	128	40	41	6970	17466
ALM2c	2	80	128	40	81	13770	34506
ALM2d	2	160	128	40	161	27370	68586
ALM3a	2	80	64	80	81	11826	22194
ALM3b	2	40	128	80	41	8610	19106
ALM3c	2	80	128	80	81	17010	37746
ALM3d	2	160	128	80	161	33810	75026
ALM4a	3	40	128	10	201	28140	79596
ALM4b	3	80	128	10	241	33740	95436
ALM5a	4	40	128	10	1641	229740	649836
ALM5b	4	40	128	10	2921	408940	1156716
ALM5c	4	80	128	10	1681	235340	665676
ALM5d	4	80	128	10	3281	459340	1299276

Table 3.7: Problem dimensions.

problem very difficult for standard optimisation approaches. For example, it makes impossible the application of Benders decomposition as discussed in Section 2.4.3.

All computations were done on the Intel Core2 Duo PC. This machine features 2 2.66GHz processors and a total of 2GB of memory.

The numerical results are collected in Table 4.3. We report the solution time, number of iterations and memory requirements for CPLEX 10.0 barrier [66] and OOPS [55, 56, 57] for each problem. Most of the problems can be solved within reasonable time and IPM iterations. Both solvers did very well for small problems. However, CPLEX ran out of memory for problems ALM5b and ALM5d, while OOPS could solve them within half an hour. For most of the problems, OOPS was faster than CPLEX, though CPLEX generally took fewer iterations. The solution time of OOPS increases steadily with the scaling of problems. When the number of assets is doubled, the solution time of OOPS increases by a factor smaller than three, which can be seen from the comparison of solution statistics of ALM1a and ALM1c, ALM2a and ALM2c, ALM3a and ALM3c. By comparing solution statistics of problems ALM1a, ALM2a and ALM3a, we can observe the influence of the number of benchmark realizations on the efficiency of both solvers. The solution statistics of ALM1b/c/d, ALM2b/c/d and ALM3b/c/d demonstrate that the solution time of CPLEX increases with the number of blocks much faster than that of OOPS. Both CPLEX and OOPS solution times are badly affected by the increase of the number of benchmark realizations. The memory requirements of OOPS are generally smaller than those of CPLEX.

Problem	CPLEX 10.0			OOPS		
	Time(s)	Itr	MEM(Mb)	Time(s)	Itr	MEM(Mb)
ALM1a	53.47	14	100.3	19.93	24	38.9
ALM1b	26.73	20	55.3	21.05	27	38.9
ALM1c	133.91	19	184.3	41.467	26	75.8
ALM1d	9.72	42	106.5	104.132	33	147.5
ALM2a	95.07	18	114.7	37.59	28	61.4
ALM2b	63.29	18	92.2	51.16	25	59.4
ALM2c	447.85	20	335.9	111.695	27	114.7
ALM2d	5021.74	35	1265.7	316.92	39	223.3
ALM3a	124.23	19	147.5	61.49	25	102.4
ALM3b	138.89	25	143.4	92.99	29	98.3
ALM3c	1072.28	30	421.9	180.91	28	190.5
ALM3d	7709.53	28	1316.9	593.562	47	376.8
ALM4a	96.89	15	196.6	72.179	28	133.1
ALM4b	588.11	15	536.6	160.20	30	262.1
ALM5a	1291.18	29	1357.8	890.44	41	1075.2
ALM5b	–	–	–	1557.15	41	1843.2
ALM5c	1542.12	20	1597.4	589.65	26	1118.2
ALM5d	–	–	–	1140.16	25	1822.7

Table 3.8: Performance Comparison of CPLEX and OOPS: solution times in seconds.

3.4.4 Parallel Solution

We also considered OOPS in parallel mode for solving large scale ALM problems with relatively symmetric scenario trees, i.e. the number of branches at the first two stages are the same. The statistics of test problems are summarized in Table 3.9. All computations were done on dual processor PCs featuring dual 3.0Ghz Intel processors running with 4GB RAM. Communication between processors is made with MPI.

Problems	Stages	Blocks	Assets	Bnmk	Total Nodes	Constraints	Variables
	T	B	I	L	$ N = \sum_{t=0}^{T-1} N_t$	$(I+L+2) N $	$(3I+L+2) N $
sALM6	3	80	64	10	6481	492556	1322124
sALM7	4	20	32	10	8421	370524	909468
sALM8	4	40	32	10	17641	776204	1905228
sALM9*	4	80	32	10	19281	848364	2082348

Table 3.9: Problem statistics of large scale models.

* 40 branches from second stage.

We report the solution statistics in terms of time consumed to solve the problems with the parallel implementation in Table 3.10. For runs on one processor, number of IPM iterations and solution times are reported. Note that sALM8 and sALM9 cannot be solved with single processor due to a lack of memory. For parallel runs with 2 and 4 processors we also report the speed-ups. We believe that the superlinear speed-up in solving problem sALM9 is due to avoiding the memory paging when more processors

Problem	1 proc		2 proc		4 proc	
	Itr	Time(s)	Time(s)	speed-up	Time(s)	speed-up
sALM6	33	2421.52	1950.51	1.24	1002.68	2.42
sALM7	52	2735.41	2560.84	1.07	1017.56	2.69
sALM8	—	—	7007.11	—	3895.60	1.80
sALM9	—	—	8077.81	—	2582.66	3.13

Table 3.10: Results of ALM problems solved in parallel mode.

are used.

Therefore, we can see that the ALM model with second-order stochastic dominance and relaxed interval second-order stochastic dominance as risk control can provide portfolio strategy with competitive performance and well-controlled underfunding. However, there exists danger that the first stage optimization is based on behavior which is not optimal once the second stage is reached, because the risk constraints consider all scenarios at the same stage together.

Chapter 4

Qualified Workforce Capacity Planning in Operational Risk Management

In the previous chapter, we dealt with allocating assets optimally. The execution of the portfolio strategy relies on the people in the back office, which involves operational risk. The human factor is an important consideration in operational risk. How to allocate labour optimally so as to manage the operational risk from the human resource aspect is the subject of this chapter. We will take the case of a back office as an example, but the result can be applied to more areas. The Stochastic Aggregate Planning Model is the main methodology of this work. The original content of most parts in this chapter has already appeared in [46], coauthored with Emmanuel Fragnière and Jacek Gondzio.

4.1 Workforce Planning Problem

In the back office of a bank, the main task is to complete the transactions of financial products and related paperwork and database management that are required. For example, a trader (front office) “writes” an over-the-counter option with a counterparty. The back office prepares the contracts, conducts all the exchange of information in due time and complies at the same time with the very stringent financial regulations. In the more and more frequent cases where the back office deals with derivative products, workers have to understand complicated pricing systems because it is part of their duties to conduct some price settlements, “reconciliations” and verifications. Several surveys (see [89, 48]) also indicate that operations in the back office are becoming far more complicated than before. The reasons for this evolution are multiple: IT harmonisation due to bank consolidation, new stringent norms and regulations affecting the operations (e.g. IAS-IFRS, Sarbannes-Oxley, Basel II, new taxation system like Qualified Intermediary), significant increase in service productivity over the last years, boom of new sophisticated (structured) financial products. So more and more complex

risks are attached to the activities in the back office and most of them require, above all, knowledge, experience and expertise to be addressed correctly and in an efficient manner.

Unfortunately, the risk management of back offices in the banking sector has not benefited from the modeling advances in other financial risks (market, credit and liquidity risks) promulgated by the Basel Committee. Basel II classifies the risks of a bank into four categories: strategic; financial; non-financial/operational; reputation and compliance risks. Regarding the non-financial/operational (risks) category, four subcategories are used: fraud, political, IT and operations (transactions mistakes, inefficiencies of processes) risks. In operations, risks have two facets: internal and external. Internal risk is typically linked to operations (hence controllable). External risks are the consequence of external causes (for instance, a change of some US GAAP accounting rules for derivatives which will affect the back office procedures as well as the training of the staff). The modeling of operational risk is rather subjective regarding its qualitative nature, being related to managerial issues.

An important aspect of the management of back offices is the notion of explicit and tacit knowledge (expertise), as explained before. Explicit knowledge enables the employees to deal with most standardised tasks. On the other hand, when a problem occurs which is not part of explicit (codified) knowledge, only the tacit knowledge (i.e. the expertise) can help to solve it. In the case of operations, the risk can “materialise” under different states [34] according to the TEID model: Threat, Event, Ignorance and Damage. Typically, the qualified worker knows how to act in these different states of risks through prevention, identification, and protective approaches.

Aggregate Planning Model (APM) is the basic prototype of the model to deal with this problem. The following extensions will be necessary.

A first improvement of the basic APM which we explore in order to better manage operational risk, is to deal with the uncertainties of capacities and demand, as will be shown in Section 4.3. In this context it corresponds to a situation where the risk is external (like a market risk) and we consider that the quality of work has no influence whatsoever on it (it is uncontrollable through managerial activities). The solution simply adapts to the evolution of the different scenarios described by the event tree.

A second improvement, as advocated by Basel II, is to calculate a probability distribution based on a Bayesian network. This reflects the observation that data related to operations risks are rarely available. We thus apply this scheme to our basic model where the demand parameters are assumed to be random. The Bayesian method, which is similar to our method in the sense of probability calculation based on Bayes rule, has been applied for example by Morton and Popova [91] to evaluate capacity strategies for a manufacturing problem.

A third improvement to deal with operational risks, and to our best knowledge implemented for the first time in a business case, is to establish a relation between the random variables and decision variables. Demand is set to be dependent on decisions.

The model becomes then far more complicated (nonlinear). It is worth mentioning that our approach is in line with the risk management requirements defined by COSO II (Committee of Sponsoring Organisation, www.coso.org) in the sense that the quality of Internal Control System affects the residual risks.

By distinguishing between explicit and implicit knowledge in our model, we will thus assume that, if more capacity of qualified workers is available, this should lead to better service because mistakes in the operations are reduced; moreover this builds up the reputation and as a consequence the demand for such service increases. For this reason, the assumption of variable and parameter independence needs to be relaxed in our stochastic programming problems.

Finally, for each developed model we use dual solutions to identify in the plan which scenarios are under stress regarding the availability of qualified workers. This analysis complements well the one provided by the primal solutions. Indeed it represents a way to price the risks due to a lack of qualified resources. We focus on the main inventory constraints of the Aggregate Planning Model, which are equality constraints with variables on both sides of the equations. In consequence, the dual solutions do not correspond to marginal values of resources. Nevertheless, dual variables (Lagrange multipliers associated with the constraints) give a relevant indication of the constraints that would need to be relaxed. For instance, they can show when liquidity should be available in addition to being able to hire additional qualified workforce capacity.

4.2 Aggregate Planning Model

In this section, we start with a general discussion of the of Aggregate Planning Model as well as its several applications. Then, how this methodology can be applied in workforce planning in operational risk management is illustrated, from the deterministic case to the stochastic case.

4.2.1 Generic Aggregate Planning Model

The Aggregate Planning Model is used to model the process of aggregation of quantities. These quantities can be products, human resources or other resources. The model aims to satisfy certain demand with the lowest cost arising from production, employment and inventories. Two existing applications of the Aggregate Planning Model are production planning and labour allocation problems.

A list of papers on production scheduling can be found in the review paper [58] which highlights the status of work in this area until 1980s, and identifies some prospective topics. The author proposed three aspects to categorise production scheduling problems. The first is by requirements generation and the problem is either open shop or closed shop, depending upon whether the requirements are generated directly by customers' orders or indirectly by inventory replenishment decisions. Although it is rare

to have a pure open or pure closed shop system, most production environments are primarily either open or closed. The second dimension refers to the number of processing steps associated with each production task or item. The processing may involve one stage or multiple stages; there may be one facility or parallel facilities which provide alternative means of processing. The third aspect indicates the measures upon which schedules are to be evaluated. One method of measurement is the scheduling cost, including production setups, overtime costs, inventory holding costs and shortage costs for not meeting deadlines or for stocking out, as well as the costs involved in generating the schedule and monitoring progress, etc. The other method of measurement is the scheduling performance, for example, utilisation level of the production resources, the percentage of late tasks, the average or maximum flow time for a set of tasks and so on.

Self-sufficiency in all activities is commonly unsatisfied in modern business enterprise, although there is high integration. There appears to be dependence between companies so that at least some work or service is provided, and subcontracting is extensively used in industry, which is “the procurement of an item or service which is normally capable of economic production in the prime contractor’s own facilities and which requires the prime contractor to make specifications available to the supplier” [72]. A subcontracting mechanism was considered as a production planning strategy in Aggregate Planning Model in [72], and a dynamic programming approach was proposed.

When a production system is composed of multiple plants, multiple products and seasonal demands, the Aggregate Planning Model with hierarchical structure may serve better, as illustrated in [63]. According to this method, the decisions are made in a sequence of four levels:

1. assignment of products to plants, using mixed-integer programming;
2. preparation of seasonal accumulation plan and, for each type of products with similar inventory costs, allocation of capacity in each plant, using linear programming;
3. detailed schedules for each product family within which products share a major setup, and allocation of capacities for families in the type, using standard inventory control methods;
4. individual planning for each item in each family, using standard inventory methods.

Extensions and modifications of this method were proposed in [114], where the model consists of three stages as the top-level plant/product-type assignment model, the family disaggregation subsystem and the item disaggregation subsystem.

Productivity change is also an important issue to be taken into account in the Aggregate Planning Model, as in [39, 78]. A learning curve is used to describe the

process of productivity changes, and this is incorporated into the Aggregate Planning Model.

An example showing how the Aggregate Planning Model could be applied to manpower planning and scheduling in the service sector is discussed in [1]. This example represents the problems occurring in most general service organisations like acute hospitals, where the demand is highly variable and must be provided in a timely manner, and where in addition, service cannot be inventoried: for instance police departments, ambulance services and fire departments. Unlike goods production which relies heavily upon the inventory capacities and backorder capacities, the service sector has to consider several important aspects. Specific labour skills should be matched with job requirements at individual work centres; the training requirements and losses in manpower efficiency that can arise from reallocation of staff; uncontrollable attrition affects the staff planning process. This work proposed a model making decisions in three levels: policy decision, staff allocation and short-term scheduling, and the model was solved using an iterative approach.

Another example of the Aggregate Planning Model applied to employment strategy is that of a telephone company [75]. The characteristic of this kind of problem is the seasonal demand for services and varying productivities of workforce which can produce various group of services. The model was formulated as a linear program, considering employment inventory constraints, production inventory constraints, the upper limit on the amount of overtime, and the upper limit on the maximum days of delay.

4.2.2 Deterministic Work Force Planning Model

Suppose we know with certainty all the parameters that are essential to make our planning decisions. In the objective function, we aim to minimise the costs resulting from employment and losses while maximising the profit. The cost function includes three elements: hiring cost, firing cost and salaries paid to employees. Qualified people have higher firing cost and salaries than non-qualified people while hiring costs are assumed to be the same for both groups. We pay only salaries to temporary employees without hiring or firing costs.

Profit is related to the volume of transactions successfully completed and hence defined as a multiple of this volume. Since we require the satisfaction of all demands on time, and since there is no inventory of transactions, this volume is actually equivalent to the demand. Finally, profit is proportional to demand. It is worth mentioning that when demand does not change in the model, i.e., it is a parameter not a variable, profit is constant and does not need to be optimised in the objective. We have the following objective function:

$$\min \sum_t ((x_h^Q)^t + (x_h^N)^t)h + (x_f^Q)^t f_Q + (x_f^N)^t f_N + (x^Q)^t w_Q + (x^N)^t w_N + (x^P)^t w_P$$

where $t = 1, \dots, T$, is the time stage (a stage corresponds to a year); x_h, x_f, x are the

numbers of people hired, fired and kept. $-Q$, $-N$ and $-P$ stand for the quantities of *Qualified* people, *Non-qualified* people and *temporary* employees, and h , f_Q , f_N , w_Q , w_N and w_P denote hiring cost, firing cost for qualified and non-qualified people, salaries per person for qualified, non-qualified and temporary employees, respectively. x^P is the number of temporary person-days, e.g. $x^P = 1$ means a temporary employee working one day. In this model, we assume that both qualified and non-qualified people are kept at work at least a full year. Hence, w_Q and w_N are both yearly salaries. Temporary people are hired daily, which means w_P is the payment to one temporary employee for working one day.

While we try to achieve the optimal value, there are two categories of constraints to be satisfied. The first group of constraints corresponds to the inventory of employees. Except for the first stage, we can hire and fire at each stage. Thus the number of employees we hold in a given period represents the number in the previous stage, minus the number fired, plus the number of newly hired employees. This is presented as follows:

$$\begin{aligned} (x^Q)^{t-1} + (x_h^Q)^t - (x_f^Q)^t &= (x^Q)^t, \quad \forall t \\ (x^N)^{t-1} + (x_h^N)^t - (x_f^N)^t &= (x^N)^t, \quad \forall t. \end{aligned}$$

The second one is the capacity constraint, which requires each demand to be completed on time at that stage. The capacities η^t at stage t are calculated in the following way:

$$(x^Q)^t \alpha^Q \times 260 + (x^N)^t \alpha^N \times 260 + (x^P)^t \alpha^P = \eta^t, \quad \forall t$$

where α^Q , α^N and α^P are work capacities for qualified, non-qualified and temporary employees, respectively. These are the numbers of transactions that one person in the corresponding group can complete per day. We assume that $\alpha^Q > \alpha^P > \alpha^N$.

We suppose that there are 260 working days per year. The capacity represents the sum of transactions processed by qualified people, non-qualified people and temporary employees in a year. The capacity must be larger than or equal to the demand β^t .

To sum up, the full mathematical programming model can be written in the following way:

$$\min \sum_t ((x_h^Q)^t + (x_h^N)^t)h + (x_f^Q)^t f_Q + (x_f^N)^t f_N + (x^Q)^t w_Q + (x^N)^t w_N + (x^P)^t w_P \quad (4.1a)$$

subject to

$$\begin{aligned} (x^Q)^{t-1} + (x_h^Q)^t - (x_f^Q)^t &= (x^Q)^t, \quad \forall t = 2, \dots, T \\ (x^N)^{t-1} + (x_h^N)^t - (x_f^N)^t &= (x^N)^t, \quad \forall t = 2, \dots, T \end{aligned} \quad (4.1b)$$

$$(x^Q)^t \alpha^Q \times 260 + (x^N)^t \alpha^N \times 260 + (x^P)^t \alpha^P = \eta^t, \quad \forall t = 1, \dots, T \quad (4.1c)$$

$$\eta^t \geq \beta^t, \quad \forall t = 1, \dots, T \quad (4.1d)$$

$$(x_h^Q)^t, (x_h^N)^t, (x_f^Q)^t, (x_f^N)^t, (x^Q)^t, (x^N)^t, (x^P)^t \in \mathbb{N}_0, \quad (4.1e)$$

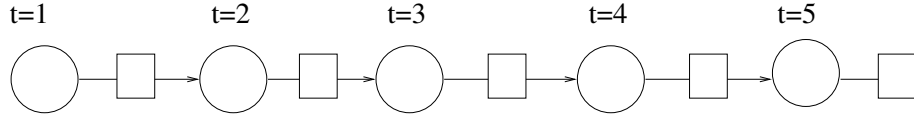


Figure 4.1: A deterministic multistage model representation: circles stand for state nodes and rectangles stand for decisions.

where $t = 1, \dots, T$ is the planning horizon and $\mathbb{N}_0 = \{0, 1, 2, 3 \dots\}$. The model above presents the standard form of the Aggregate Planning Model, with (4.1a) as the cost objective function to be minimised, (4.1b) corresponding to the dynamic constraints and (4.1c), (4.1d) being the local constraints.

Figure 4.1 reflects a 5-period instance ($T=5$) [45]. Decision variables are the number of people to be employed. Consequently they are defined as integers, which means that the model is an integer programming problem. In reality, however, uncertainty needs to be taken into account when planning. We will develop the stochastic Aggregate Planning Model in the following section.

4.3 Multi-stage Stochastic Aggregate Planning Model of Workforce Planning Problem

In this section, we develop stochastic programs with randomness in demand and capacities.

4.3.1 Random Demand Parameters

We can assume that the demands are random parameters and can thus take several possible values. This is usually modelled as an event tree (see for instance Figure 4.2). As a result, we obtain a multistage stochastic program.

The deterministic equivalent formulation of our Aggregate Planning Model with uncertain demands becomes:

$$\min \sum_{t, j_t} P_{j_t}^t (((x_h^Q)_{j_t}^t + (x_h^N)_{j_t}^t)h + (x_f^Q)_{j_t}^t f_Q + (x_f^N)_{j_t}^t f_N + (x_Q)_{j_t}^t w_Q + (x^N)_{j_t}^t w_N + (x^P)_{j_t}^t w_P) \quad (4.2a)$$

subject to

$$(x^Q)_{a(t, j_t)}^{t-1} + (x_h^Q)_{j_t}^t - (x_f^Q)_{j_t}^t = (x^Q)_{j_t}^t, \quad \forall t = 2, \dots, T, \quad \forall j_t \in N_t \quad (4.2b)$$

$$(x^N)_{a(t, j_t)}^{t-1} + (x_h^N)_{j_t}^t - (x_f^N)_{j_t}^t = (x^N)_{j_t}^t, \quad \forall t = 2, \dots, T, \quad \forall j_t \in N_t, \quad (4.2c)$$

$$(x^Q)_{j_t}^t \alpha^Q \times 260 + (x^N)_{j_t}^t \alpha^N \times 260 + (x^P)_{j_t}^t \alpha^P = \eta_{j_t}^t, \quad \forall t = 1, \dots, T, \quad \forall j_t \in N_t, \quad (4.2d)$$

$$\eta_{j_t}^t \geq \beta_{j_t}^t, \quad \forall t = 1, \dots, T, \quad \forall j_t \in N_t, \quad (4.2e)$$

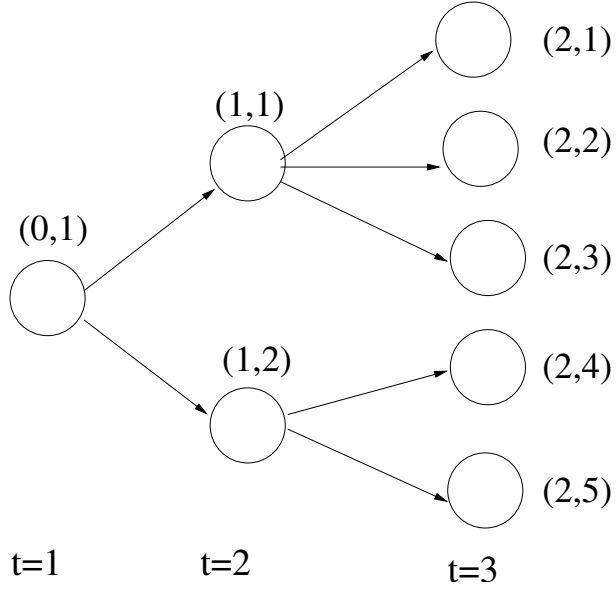


Figure 4.2: An example of an event tree describing different demand states of nature.

$$(x_h^Q)_{j_t}^t, (x_h^N)_{j_t}^t, (x_f^Q)_{j_t}^t, (x_f^N)_{j_t}^t, (x^Q)_{j_t}^t, (x^N)_{j_t}^t, (x^P)_{j_t}^t \in \mathbb{N}_0,$$

where $j_t \in N_t = \{1, \dots, n_t\}$, is the set of demand states, and P is the probability distribution of demand, which defines (partial) path probabilities: $P_{j_t}^t$ is the probability (at the start) that a path goes through node j_t at time t and $a(t, j_t)$ denotes the ancestor of node j_t in the event tree. This formulation remains an integer programming problem.

4.3.2 Random Capacity Parameters

Due to some unpredictable events, like illness, holiday, or some unexpected accidents, constant capacity of employees cannot be guaranteed. Hence, we introduce uncertain capacity parameters into the model, $\tilde{\alpha}^Q$ and $\tilde{\alpha}^N$. Temporary employees are just employed in cases where we need more people. Consequently, we assume that they only take a small part of the total demand and thus we neglect the variation of their capacities. The corresponding capacity constraints become:

$$(x^Q)_{j_t}^t \alpha_{j_t}^Q \times 260 + (x^N)_{j_t}^t \alpha_{j_t}^N \times 260 + (x^P)_{j_t}^t \alpha^P = \eta_{j_t}^t, \quad (4.3)$$

$$\eta_{j_t}^t \geq \beta_{j_t}^t, \quad (4.4)$$

where $t = 1, \dots, T$, $j_t = 1, \dots, n_t$.

The other concern about capacity is new employees' capability. There is always a certain period needed for a person who is newly employed to get used to the work environment and to become familiar with their responsibility. We cannot expect a new employee to be as efficient as an experienced one. Hence, their capabilities have to be

valued separately. This is reflected in the following capacity constraint:

$$\begin{aligned} ((x^Q)_{j_t}^t - (1 - \delta_{j_t})(x_h^Q)_{j_t}^t)\alpha_{j_t}^Q \times 260 &+ ((x^N)_{j_t}^t - (1 - \delta_{j_t})(x_h^N)_{j_t}^t)\alpha_{j_t}^N \times 260 \\ &+ (x^P)_{j_t}^t \alpha^P = \eta_{j_t}^t, \end{aligned} \quad (4.5)$$

where $t = 1, \dots, T$, $j_t = 1, \dots, n_t$, and $0 < \delta < 1$ is the ratio of a new employee's capability to an experienced employee's capability, i.e., when an experienced employee completes one transaction, the new employee can do only δ transactions. For a new hired employee, we have to subtract the lack of capacity $(1 - \delta_{j_t}) \times (x_h^Q)_{j_t}^t$ and $(1 - \delta_{j_t}) \times (x_h^N)_{j_t}^t$ from the total capacity. After one year, the new employee can work as efficiently as other employees.

The randomness in demand and capacity in the present model are both external risks which are not controlled by the decisions in the model. In Section 4.5, we will introduce the internal risk from the uncertain demand that depends on the decisions.

4.4 Revising Operation Efficiency Probability Distributions

Operation efficiency is a way to measure the work done by employees, which is how many transactions an employee completes per unit time (labour cost) and how many errors an employee makes per transaction (error rate). In banking, a mistake in operations could bring big losses to the company and also decrease the demand in subsequent stages. Limiting the number of mistakes to the strict minimum is essential. This notion of operational efficiency is thus intimately linked to operational risk management [27, 47], and is related to people's knowledge and skills. Even if we know whether people are qualified or not, errors can still happen unexpectedly. Hence it is necessary to consider randomness of operation efficiency in the model. And we need the probability distribution describing the behaviour of the random variable. However, in operational risk management, this kind of data is hard to collect. Firstly, long-term data is lacking, which means only the data for a few recent years is available. Secondly, a company never publishes its errors and operational losses and this makes the data unavailable. It is indeed extremely difficult to collect sufficient years of statistics describing operational risks, so as to be then able to assume any theoretical behaviour. Because of this, we will use a Bayesian Network.

Bayesian Networks (BN) have emerged as a method of choice to deal with operational risks, especially in the banking sector, because their use does not necessitate the gathering of huge amounts of past data. BN is in fact grounded upon classical decision theory and also adopts computing schemes of Artificial Intelligence [67]. Typically with a Bayesian approach we start with a subjective probability associated with a particular event. The a priori probability (also called subjective probability as opposed to objective probability) is assessed by the manager and corresponds to their own intuition and

expertise. Along the way, obtaining new imperfect information and depending upon the quality of past imperfect information provided by the issuer, the manager will be more or less inclined to modify his initial judgement. In a formal model this would be called the a posteriori analysis where a priori probabilities are modified using the Bayes formula. The Bayesian Network can be presented as a quantitative approach to handling qualitative dynamic choices. On the other hand, through the modelling of decision trees this approach should enable the manager to structure the dynamic dimension of the decision process.

4.4.1 Bayesian Network

The Bayesian method is a statistical inductive way to update the probability distribution using Bayes' theorem which was first stated by Bayes in the 18th century. It makes full use of the available information and data to achieve a more reliable probability distribution of the uncertainties. This theory [10] has been widely applied in several areas including economics, informatics, biostatistics, educational and psychological research, social science, decision theory and optimisation, etc. The application of Bayesian method to Weibull process can be found in [118], which induced the posterior distribution from the prior distribution, where both the shape parameter and the scale parameter are unknown, and are given a discrete distribution and a gamma distribution, respectively. Two examples, namely the reliability of communication satellites, and also the replacement of fuel pumps, were explained using the Bayesian method. The Weibull process as a natural conjugate prior is quite popular in the Bayesian method. In [8] the author discussed how to achieve the minimal cost with the optimal overhaul interval when the restoration is Weibull distributed. It was also proved experimentally that the Bayesian method is superior to the Naive T model, and that its performance can be significantly improved by eliminating the bias in the prior estimates. Another research in this topic [19] described the maintenance policies in both the deterministic case and the stochastic case with the number of minimal repairs modelled as a Weibull process during the warranty period. In the stochastic case, two Bayesian policies are provided when the failure parameters of the Weibull process are unknown. While one updates the policy at each renewal point, the other one does so at both failure and renewal time. The simulation results have shown that, if either the failure is overestimated or underestimated, Bayesian policies could decrease the costs. Again, improved knowledge of the prior can lead to better results.

In [131], the Bayesian method was proved to be one of the most efficient ways in terms of information processing, which means that the new data is 100% transformed into the prior distribution to get the posterior distribution. While more contributions were made to the Bayesian analysis, in [70] it was chosen to develop the technology for implementations used by experimenters in multiple normal linear regression models. In this paper, the interactive elicitation was explored to build up the prior distribution rather than the experimental data.

Due to the nature of Bayesian method, it is widely used in stochastic scheduling. The dynamic allocation index theorem can provide the generalised optimal strategy for a family of alternative bandit processes, where the parameters of the random elements have a prior distribution, in terms of lowest cost or least time [49]. The problem of job allocation with a single machine was discussed in [59]. All the jobs in different families have random processing time whose parameters are random as well. The Bayesian method was applied to update the distribution of these random parameters. The paper started with the simplified situations, i.e. where jobs in the same family are processed successively, and in which there are two classes of jobs, one of which has a known parameter. The author then solved the problem as a general m-class problem. Besides the job allocation problem, another scheduling problem is concerned with manpower. Papers [91, 105] by Morton and Popova researched the manpower scheduling problem when the machine downtime and working hours are randomly distributed with unknown parameters. Mathematical programming is the main methodology to solve this problem. The model aimed to minimise the penalties from both late deliveries and exceeding target budget in the objective under the condition that the working cannot be larger than the machine available hours. The demand was allowed to be unsatisfied with penalty and the production cost was equal to wages paid for regular-time working hours and overtime. In the deterministic model, Bayesian estimation provides the point forecasts for the up-hours and production per shift for each shaft type, while in the stochastic situation, the distributions of the up-hours and production rate are estimated by the Bayesian method. The computational results show that, compared to empirical distributions, the expected costs of the schedule are much lower when generated by the Bayesian predictive distribution.

The Bayesian Network (BN) has been widely applied as a successful description of causalities in several areas such as diagnosis, heuristic search, ecology, data mining and intelligent trouble shooting systems. It is defined as follows [67]:

Definition A Bayesian network consists of the following:

- A set of variables and a set of directed edges between variables.
- Each variable has a finite set of mutually exclusive states.
- The variables together with the directed edges form a Directed Acyclic Graph(DAG). (A directed graph is acyclic if there is no directed path $A_1 \rightarrow \cdots \rightarrow A_n$ such that $A_1 = A_n$, where A_i are variables.)
- To each variable A with parents B_1, \cdots, B_n , there is attached the potential table $P(A|B_1, \cdots, B_n)$. A is a child of B and B is a parent of A , if there is a link directed from B to A . And, always, the parent(s) are set to be cause(s) of the child.

$P(A|B_1, \cdots, B_n)$ is the conditional probability of A given (B_1, \cdots, B_n) . Note that if A has no parents then the conditional probability reduces to unconditional probabilities

$P(A)$. Causal relations also have a quantitative side, namely their strength. This is expressed by attaching numbers to the relation links. In BN, it is natural to set the conditional probability to be the strength of the link. Let B be a parent of A , then $P(A|B)$ is the strength of their link.

The two fundamental rules for probability calculus are

$$\begin{aligned} P(A|B)P(B) &= P(A, B), \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)}, \end{aligned}$$

where $P(A, B)$ is the probability of the joint event that both A and B happen. Sometimes $P(A|B)$ is called the likelihood of B given A . Assume A has n outcomes $a_1 \cdots a_n$, with an effect on the event B , where B is known. Then, $P(a_i|B)$ is a measure of how likely it is that a_i is the cause. In particular, if all a_i 's have prior probabilities, Bayes' rule yields

$$P(a_i|B) = \frac{P(B|a_i)P(a_i)}{P(B)}.$$

How to calculate the probabilities in a Bayesian network is given in the following theorem:

Theorem 2 (The chain rule, [67]). *Let $U = (A_1, \dots, A_n)$ be a universe of variables and define a Bayesian network over this set. Then the joint probability distribution $P(U)$ is the product of all conditional probabilities specified in the Bayesian network:*

$$P(U) = \prod_i P(A_i | pa(A_i)),$$

where $pa(A_i)$ is the parent set of A_i .

Let A be a variable in a Bayesian network, with prior probability distribution $P(A) = (p_1, \dots, p_n)$. Assume we get the information (or evidence) e that A only has two possible states i and j , i.e. the belief becomes $P(A, e) = (0, \dots, 0, p_i, 0, \dots, 0, p_j, 0, \dots, 0)$. Note that the prior probability of this information is $P(e) = p_i + p_j$, the sum of the probabilities of the possible states. Using the fundamental rule, we have:

$$P(A|e) = \frac{P(A, e)}{P(e)} = \frac{P(A, e)}{\sum_A P(A, e)}.$$

Here, we can interpret $P(A, e)$ as a multiplication of the prior probability distribution $P(A)$ with the vector $\underline{e} = (0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0)$. This n -dimensional binary vector is called a finding on A .

Now, we extend this probability update to the whole variable universe. Given U as a universe of variables, we have a finding as above, which results in the belief $P(U, e)$ from the prior probability distribution $P(U)$. Similarly, $P(e)$ is the sum of all entries

in $P(U, e)$, $P(U, e)$ is the product of $P(U)$ with the finding \underline{e} and

$$P(U|e) = \frac{P(U, e)}{P(e)} = \frac{P(U, e)}{\sum_U P(U, e)}.$$

If e consists of m findings $\underline{f}_1, \dots, \underline{f}_m$ which are m binary vectors, then the probability can be updated according to the theorem below:

Theorem 3 ([67]). *Let U be a universe of variables and let $e = \{\underline{f}_1, \dots, \underline{f}_m\}$. Then*

$$P(U, e) = P(U) \times \underline{f}_1 \times \dots \times \underline{f}_m \text{ and } P(U|e) = \frac{P(U, e)}{P(e)}$$

where

$$P(e) = \sum_U P(U, e).$$

An example of a Bayesian network is shown in Figure 4.3, assuming probability distribution $P(U)$ and that a (corresponding) conditional probability distribution for each arrow is defined. The dotted lines with e indicate insertion of evidence, which can be used to update the probability distribution.

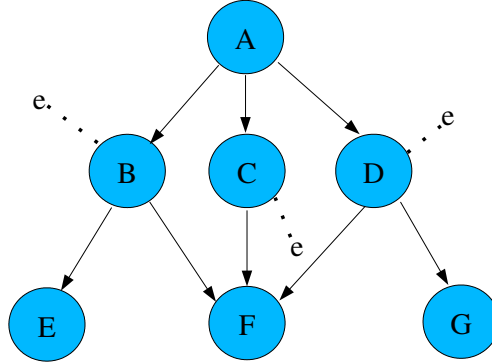


Figure 4.3: An example of a Bayesian Network.

An example from operational risk management. Consider the settlement process of trades in a bank. The trade volume (TC) could influence the trade capture (TC) in front office and the payment input in back office. Those two together will influence the settlement (S) result. Meanwhile, the trade type (TT) which could be simple or complicated has an effect on trade match confirmation (CM). And this match

can also affect the settlement (S) result. A Bayesian network presenting these cause-effect relations is shown in Fig 4.4. The probability distribution of the nodes without any parent and conditional probability distribution of other nodes are given in Table 4.4.1

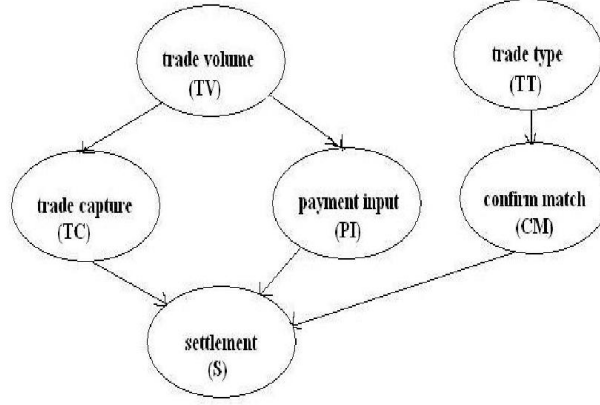


Figure 4.4: A Bayesian network presenting settlement process in a bank.

Thus, the universe of variables is $U = (TV, TT, TC, PI, CM, S)$. Following Theorem 2, we can calculate the probability distribution of U , e.g.

$$\begin{aligned}
& P(U = (high, complex, correct, correct, correct, succeed)) \\
&= P(TV = high) \times P(TT = complex) \times P(TC = correct|TV = high) \\
&\times P(PI = correct|TV = high) \times P(CM = correct|TT = complex) \\
&\times P(S = succeed|TC = correct) \times P(S = succeed|PI = correct) \\
&\times P(S = succeed|CM = correct) \\
&= 0.4 \times 0.3 \times 0.9 \times 0.9 \times 0.9 \times 0.95 \times 0.95 \times 0.95 \\
&= 0.075,
\end{aligned}$$

which is the probability of the case that the trade volume is high, trades are in simple type and captured correctly, no error is in the payment input, match confirmation is correct, and the final settlement succeeds; similarly,

$$\begin{aligned}
& P(U = (low, simple, correct, correct, correct, succeed)) \\
&= P(TV = low) \times P(TT = simple) \times P(TC = correct|TV = low) \\
&\times P(PI = correct|TV = low) \times P(CM = correct|TT = simple) \\
&\times P(S = succeed|TC = correct) \times P(S = succeed|PI = correct) \\
&\times P(S = succeed|CM = correct) \\
&= 0.6 \times 0.7 \times 0.99 \times 0.99 \times 0.99 \times 0.95 \times 0.95 \times 0.95 \\
&= 0.349.
\end{aligned}$$

TV	P(TV)	TT	P(TT)
high	0.4	simple	0.7
low	0.6	complex	0.3

(a) trade volume

(b) trade type

TC	P(TC TV=high)	P(TC TV=low)	PI	P(PI TV=high)	P(PI TV=low)
correct	0.9	0.99	correct	0.9	0.99
error	0.1	0.01	error	0.1	0.01

(c) trade capture

(d) payment input

CM	P(CM TT=simple)	P(CM TT=complex)
correct	0.9	0.99
error	0.9	0.99

(e) confirm match

S	P(S TC=c)	P(S TC=e)	P(S PI=c)	P(S PI=e)	P(S CM=c)	P(S CM=e)
succeed	0.95	0.01	0.95	0.01	0.95	0.01
fail	0.05	0.99	0.05	0.99	0.05	0.99

(f) settlement

Table 4.1: Probability distribution and conditional probability distribution of the operational risk management Bayesian Network.

"c" for "correct"; "e" for "error".

Suppose now we know that the trade is captured correctly (i.e. $TC=c$), which is an evidence . By Theorem 3, the probability distribution of U can be revised, e.g.

$$\begin{aligned}
& P(U = (high, complex, correct, correct, correct, succeed)|TC = c) \\
&= \frac{P((high, complex, correct, correct, correct, succeed), TC = c)}{P(TC = c)} \\
&= \frac{0.075}{0.954} \\
&= 0.079 \\
& P(U = (low, simple, correct, correct, correct, succeed)|TC = c) \\
&= \frac{P((low, simple, correct, correct, correct, succeed), TC = c)}{P(TC = c)} \\
&= \frac{0.349}{0.954} \\
&= 0.366
\end{aligned}$$

The tree presented in Figure 4.5 shows how the probability of this Bayesian network is revised. Initially the priori probabilities are assigned to the root nodes and conditional probabilities are assigned to each edge. By collecting information (or evidence) of the variables, which are observable, the probability of the whole network can be updated.

□

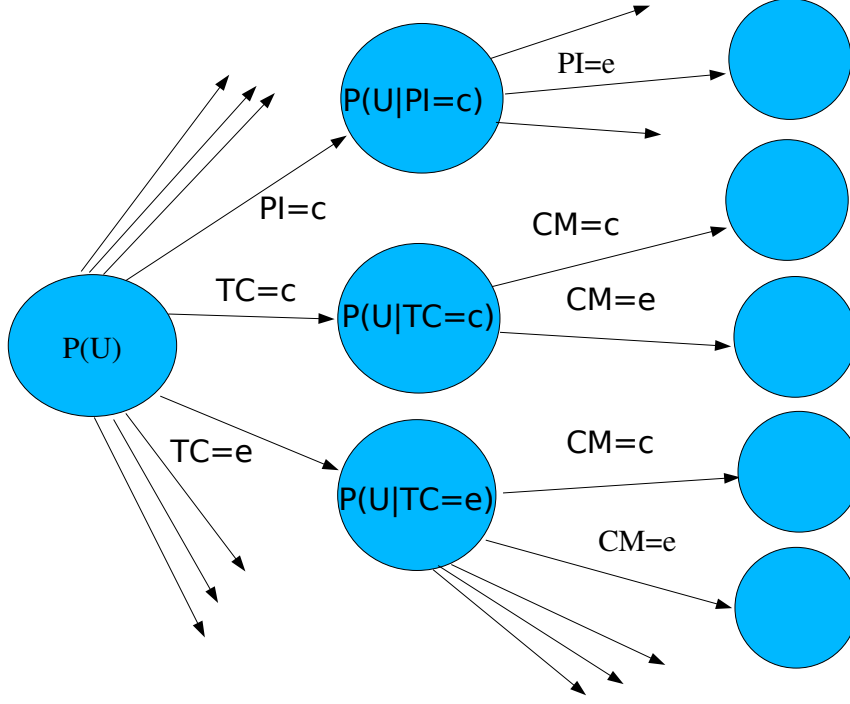


Figure 4.5: Probability revision tree of the Bayesian network.

4.4.2 Probability Distribution Revision

In operational risk, unexpected variability in operation efficiency has a significant effect on losses. Applying BN in this context means using the information on losses presently available to revise the probability distribution of operation efficiency. We illustrate the use of BN in an example below.

Let Er^Q be the operation efficiency of qualified people, which is the expected number of errors made by qualified people in one transaction, and similarly, let Er^N be the operation efficiency of non-qualified employees. One error will lead to a failure of the current transaction. We assume that $0 \leq Er^Q \leq Er^N \leq 1$. Firstly we need to know an a priori probability of the operation efficiency from experts' knowledge. For example, we denote A as $Er^Q \geq \phi$ which means qualified employees have shown (a) minimal (standard of) efficiency and \bar{A} as $Er^Q < \phi$, where ϕ is a benchmark set to judge the operation efficiency. Suppose that the experts give the following estimates: $P(A) = 0.2$, $P(\bar{A}) = 0.8$. In addition, denote B as $Loss > 0$ which is the loss resulted from the activity of qualified people, and \bar{B} as $Loss = 0$. Again, suppose that experts also determine that $P(B|A) = 0.7$ and $P(B|\bar{A}) = 0.1$. We can calculate

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = 0.22. \quad (4.6)$$

Now suppose that after one year, we know that a loss has occurred, which is the

event B. We get new confidence about the operation efficiency by BN as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = 0.64. \quad (4.7)$$

In this simple example we notice that the probability of poor operation efficiency increases from 0.2 to 0.64. For the random variables with more than 2 possible values, we still need to revise all the probabilities in the same way with (4.7). In addition, after revising the probabilities of A , we also need to update B 's probabilities following the same approach as (4.6).

We can check that the revision process corresponds to our intuition. The conditional probability of positive loss given poor operation efficiency is higher than the unconditional probability when we know nothing about the operation efficiency, which is

$$P(B|A) \geq P(B). \quad (4.8)$$

Then, when loss happens, we revise A 's probability as follows:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \geq P(A). \quad (4.9)$$

The inequality (4.9) follows easily from (4.8). This revision tells us that incurring loss increases the probability of poor operation efficiency, which agrees with people's intuitive judgement.

Finally, integrating the BN framework to revise the probability distribution of operation efficiency, in the Aggregate Planning Model, we obtain the following equations:

$$\begin{aligned} \min \quad & \sum_{t,j_t} P_{j_t}^t (((x_h^Q)_{j_t}^t + (x_h^N)_{j_t}^t)h + (x_f^Q)_{j_t}^t f_Q + (x_f^N)_{j_t}^t f_N \\ & + (x^Q)_{j_t}^t w_Q + (x^N)_{j_t}^t w_N + (x^P)_{j_t}^t w_P \\ & + ((x^Q)_{j_t}^t \alpha_{j_t}^Q Er_{j_t}^Q + (x^N)_{j_t}^t \alpha_{j_t}^N Er_{j_t}^N) \times 260 \times \theta \end{aligned} \quad (4.10a)$$

subject to

$$\begin{aligned} (x^Q)_{a(t,j_t)}^{t-1} + (x_h^Q)_{j_t}^t - (x_f^Q)_{j_t}^t &= (x^Q)_{j_t}^t, \quad j_t = 1, \dots, N_t, \quad \forall t = 2, \dots, T, \\ (x^N)_{a(t,j_t)}^{t-1} + (x_h^N)_{j_t}^t - (x_f^N)_{j_t}^t &= (x^N)_{j_t}^t, \quad j_t = 1, \dots, N_t, \quad \forall t = 2, \dots, T, \end{aligned} \quad (4.10b)$$

$$\begin{aligned} ((x^Q)_{j_t}^t - (1 - \delta_v)(x_h^Q)_{j_t}^t) \alpha_{j_t}^Q \times 260 &+ ((x^N)_{j_t}^t - (1 - \delta_v)(x_h^N)_{j_t}^t) \alpha_{j_t}^N \times 260 \\ &+ (x^P)_{j_t}^t \alpha^P = \eta_{j_t}^t, \end{aligned} \quad (4.10c)$$

$$j_t = 1, \dots, N_t, \quad \forall t = 1, \dots, T,$$

$$\eta_{j_t}^t \geq \beta_{j_t}^t, \quad (4.10d)$$

$$j_t = 1, \dots, N_t, \quad \forall t = 1, \dots, T,$$

$$(x_h^Q)_{j_t}^t, (x_h^N)_{j_t}^t, (x_f^Q)_{j_t}^t, (x_f^N)_{j_t}^t, (x^Q)_{j_t}^t, (x^N)_{j_t}^t, (x^P)_{j_t}^t \in \mathbb{N}_0,$$

where the probability distributions are calculated according to the following equations:

$$P_{j_t}^t = \frac{P_{a(t,j_t)}^{t-1} P_{loss}(Loss > 0 | j_t)}{P_{loss}^{t-1}(Loss > 0)} , \text{ if loss happens,} \quad (4.11a)$$

$$j_t = 1, \dots, n_t, \forall t = 2, \dots, T,$$

$$P_{j_t}^t = \frac{P_{a(t,j_t)}^{t-1} (1 - P_{loss}(Loss > 0 | j_t))}{1 - P_{loss}^{t-1}(Loss > 0)} , \text{ if no loss happens,} \quad (4.11b)$$

$$j_t = 1, \dots, n_t, \forall t = 2, \dots, T,$$

$$P_{loss}^t(Loss > 0) = \sum_{j_t} P_{loss}(Loss > 0 | j_t) P_{a(t,j_t)}^{t-1} , \quad (4.11c)$$

$$j_t = 1, \dots, n_t, \forall t = 2, \dots, T.$$

θ is the loss in capital per error. (4.11a), (4.11b) and (4.11c), impose the Bayesian Network, where $P_{j_t}^t$ denotes the probability distribution of operation efficiency, P_{loss} is the probability distribution of the loss while $P_{loss}(\cdot)$ is the conditional probability distribution of loss conditioned on the operation efficiency. (4.11a) and (4.11b) are the probability revisions of operation efficiency. At each stage t , we revise the probabilities from previous stage $t - 1$ by collecting the *loss* information. The probability distribution revised at each stage is essential in the objective function and clearly influences the decision-making accordingly. (4.11c) calculates the new probability of loss after operation efficiency probability revision at each stage. See Figure 4.6. The probability is revised at every node after knowing loss state of the corresponding stage, e.g. at $(t - 1, a(j_t))$ loss state of stage t is unknown; after collecting the state information at t , the probability is revised at (t, j_t) . As time progresses, we continuously collect the loss state information to update the probability distributions. The optimisation problem (4.10) is a linear integer program.

4.5 Random Parameters Dependent on Decisions

Demand can be dependent upon several factors, such as the trend of the market or the management of the company. One of the most important factors is the reputation acquired based on the quality of transactions processed so far, which is essentially related to individual employee expertise. When customers receive products or services of high quality, they are more likely to continue the business and even increase the volume and bring more business to the company, which will increase the demand at the next stage. Conversely, if customers are not satisfied with what they have got, they may change to other companies.

The operational risk in this case has two issues. Firstly, demand may decrease due to the low quality of service provided and this would result in revenue drop. According to COSO philosophy (www.coso.org), the crude risk is reduced as a function of the

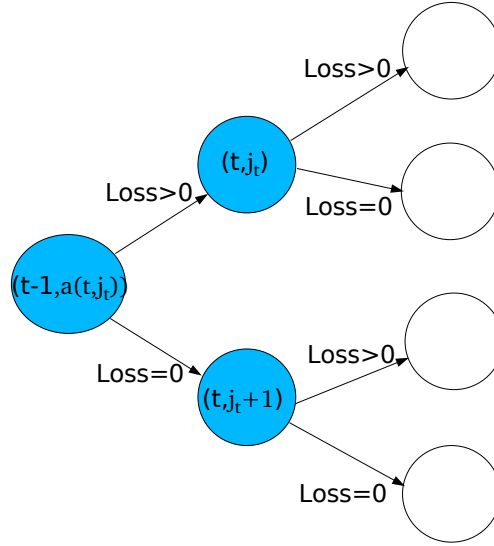


Figure 4.6: The process of probability distribution revision, where $a(t, j_t)$ is the ancestor node of j_t .

level of quality of the internal control system. As an example, we can imagine that a bank's main activity is in the development and trading of structured products (the crude risk is thus huge in terms of financial and operational risks). However, if highly qualified people are dealing with these activities, the internal control system presents a high level of quality, due to the fact that the qualitative skills like information search style, level of education and training on risk, influence the capability of risk managers to identify risks. Consequently, the residual risk is minimised. The managerial treatment of risk becomes thus crucial in regard to the COSO philosophy. Another element that has to be taken into account in the management of back offices is the distinction of explicit and tacit knowledge, as discussed before. Explicit knowledge is easy to learn and follow. They can be used to deal with the routine tasks which obey standards and can be documented in working procedures. In contrast, tacit knowledge needs much more effort to express, formalise or convey, resulting in difficulties to master this knowledge and to make good use of it. However, tacit knowledge is necessary to tackle the unexpected problems affecting the service production that cannot be documented, when explicit knowledge can rarely make a contribution. In such situation, qualified employees with both tacit and explicit knowledge are expected to drive it back to normal. Hence, we assume that if more qualified workers are available, this should lead to better service.

Our modified Aggregate Planning Model will indicate how many of the qualified and non-qualified people should be employed, which leads to the second issue concerning the risk. If demand increases because of excellent services received by customers and good reputations built up, the lack of workforce becomes a risk. Such demand growth

is largely determined by the decisions of the model, i.e. volume of qualified people, and cannot be captured by statistical behaviour prediction. It is thus essential to take into account the relations of demand and decisions in the model in order to properly manage the risk of lack of expertise. This also illustrates the problem that operational risk in this perspective cannot be treated in the same way as market risk. Indeed, market risks result from market fluctuations which are by definition non-controllable. This is not the case for operational risks whose origins are human and can, to a certain extent, often be controlled internally.

Operation efficiency is used to measure the work done by employees. Qualified people have additional professional knowledge and skills enabling them to achieve a higher throughput with a lower rate of error than their non-qualified colleagues. However, on the other hand, non-qualified employees are much cheaper to employ in both salary and firing terms. We attempt to determine the number of qualified people required to minimise losses due to employee error, hence not only directly impacting profits, but also growing demand in subsequent stages as a result of better customer experience. We are therefore trading off the additional cost of qualified employees against the reduction in error-based losses and the growth in demand they produce.

To measure employees' work in a macro view, we use the total number of errors that employees make in a year, denoted by $\#E$:

$$\#E = ((x^Q)\alpha^Q Er^Q + (x^N)\alpha^N Er^N) \times 260.$$

Since the reputation could be ruined by errors, one error would lead to a reduction of more than one unit of demand. The decrease of demand due to errors can be:

$$\beta_{dec} = (\#E)^\tau,$$

where $1 \leq \tau \leq 2$. Meanwhile, those transactions done correctly, denoted by $\#C$, can bring more business, which means the demand can increase by the following amount:

$$\beta_{inc} = \lambda \#C,$$

where $0 < \lambda < 1$. Demand increase that is induced by transactions completed correctly is less than the decrease due to the same amount of transactions with errors. This is consistent with the fact that the reputation is much easier destroyed than built up.

When operation efficiency is not considered, demand can be estimated according to the market and other conditions, denoted by $\tilde{\beta}$. Due to operation efficiency, the value of demand could differ from $\tilde{\beta}$ in the following way:

$$\beta = \tilde{\beta} + \beta_{inc} - \beta_{dec} = \tilde{\beta} + \lambda \#C - \#E^\tau,$$

where $0 < \lambda < 1$, $1 \leq \tau \leq 2$. Nonlinearity is introduced by this function. This

polynomial function, however, can approximately be linearised by a Special Order Set 2 (SOS2) [124] as described below.

For a pair of variables (β_{dec}, E) where $\beta_{dec} = E^\tau$, τ is a real number and $E \in \mathbb{R}^+$, we can calculate m pairs of values as (β_{dec}^i, E_i) , for $i = 1, \dots, m$. Then, given a value of E denoted as E^p , the corresponding value of β_{dec} can be approximated as

$$\begin{cases} \beta_{dec}^p &= \sum_{i=1}^m \rho_i \beta_{dec}^i, \\ E^p &= \sum_{i=1}^m \rho_i E_i, \\ 1 &= \sum_{i=1}^m \rho_i \end{cases}$$

The nonlinear function now becomes a series of linear functions.

Demand changes with decisions, which makes the profit dependent on decisions too, and therefore we take it into account in the objective function: it contributes a term $\beta^t \cdot \gamma$, where γ is the income per transaction completed. Based on the model (4.10), by adding the demand changing function as (4.12e), (4.12f) and (4.12g), the model is presented as follows:

$$\begin{aligned} \min \quad & \sum_{t,j_t} P_{j_t}^t ((x_h^Q)_{j_t}^t + (x_h^N)_{j_t}^t)h + (x_f^Q)_{j_t}^t f_Q + (x_f^N)_{j_t}^t f_N \\ & + (x^Q)_{j_t}^t w_Q + (x^N)_{j_t}^t w_N + (x^P)_{j_t}^t w_P \\ & + ((x^Q)_{j_t}^t \alpha_{j_t}^Q Er_{j_t}^Q + (x^N)_{j_t}^t \alpha_{j_t}^N Er_{j_t}^N) \times 260 \times \theta - \beta_{j_t}^t \gamma \end{aligned} \quad (4.12a)$$

subject to

$$\begin{aligned} (x^Q)_{a(t,j_t)}^{t-1} + (x_h^Q)_{j_t} - (x_f^Q)_{j_t}^t &= (x^Q)_{j_t}^t, \quad j_t = 1, \dots, N_t, \quad \forall t = 2, \dots, T, \\ (x^N)_{a(t,j_t)}^{t-1} + (x_h^N)_{j_t} - (x_f^N)_{j_t}^t &= (x^N)_{j_t}^t, \quad j_t = 1, \dots, N_t, \quad \forall t = 2, \dots, T, \end{aligned} \quad (4.12b)$$

$$\begin{aligned} ((x^Q)_{j_t}^t - (1 - \delta_{j_t})(x_h^Q)_{j_t}^t) \alpha_{j_t}^Q \times 260 &+ ((x^N)_{j_t}^t - (1 - \delta_{j_t})(x_h^N)_{j_t}^t) \alpha_{j_t}^N \times 260 \\ &+ (x^P)_{j_t}^t \alpha^P = \eta_{j_t}^t, \end{aligned} \quad (4.12c)$$

$$j_t = 1, \dots, N_t, \quad \forall t = 1, \dots, T,$$

$$\eta_{j_t}^t \geq \beta_{j_t}^t, \quad j_t = 1, \dots, N_t, \quad \forall t = 1, \dots, T, \quad (4.12d)$$

$$\#E_{j_t}^t = ((x^Q)_{j_t}^t \alpha_{j_t}^Q Er_{j_t}^Q + (x^N)_{j_t}^t \alpha_{j_t}^N Er_{j_t}^N) \times 260, \quad j_t = 1, \dots, N_t, \quad \forall t = 1, \dots, T, \quad (4.12e)$$

$$\#C_{j_t}^t = \beta_{j_t}^t - \#E_{j_t}^t, \quad j_t = 1, \dots, N_t, \quad \forall t = 1, \dots, T, \quad (4.12f)$$

$$\beta_{j_{t+1}}^{t+1} = \tilde{\beta}_{j_{t+1}}^{t+1} + \lambda \#C_{a(t,j_t)}^t - (\#E_{a(t,j_t)}^t)^\tau, \quad j_t = 1, \dots, N_t, \quad \forall t = 1, \dots, T-1, \quad (4.12g)$$

$$(x_h^Q)_{j_t}^t, (x_h^N)_{j_t}^t, (x_f^Q)_{j_t}^t, (x_f^N)_{j_t}^t, (x^Q)_{j_t}^t, (x^N)_{j_t}^t, (x^P)_{j_t}^t \in \mathbb{N}_0,$$

where the probability distributions are revised in the following way:

$$P_{j_t}^t = \frac{P_{a(t,j_t)}^{t-1} P_{loss}(Loss > 0 | j_t)}{P_{loss}^{t-1}(Loss > 0)}, \quad \text{if loss happens,} \quad (4.13a)$$

$$j_t = 1, \dots, n_t, \quad \forall t = 2, \dots, T,$$

$$P_{j_t}^t = \frac{P_{a(t,j_t)}^{t-1}(1 - P_{loss}(Loss > 0|j_t))}{1 - P_{loss}^{t-1}(Loss > 0)}, \text{ if no loss happens,} \quad (4.13b)$$

$$j_t = 1, \dots, n_t, \forall t = 2, \dots, T,$$

$$P_{loss}^t(Loss > 0) = \sum_{j_t} P_{loss}(Loss > 0|j_t) P_{a(t,j_t)}^{t-1}, \quad (4.13c)$$

$$j_t = 1, \dots, n_t, \forall t = 2, \dots, T.$$

Since the nonlinear demand function (4.12g) can be linearised, this model still can be solved by integer solvers.

It is worth mentioning that while demand is influenced by previous decisions, there exists also an influence in the opposite direction. β^t is one of the main factors affecting the decision at stage t . Conversely the decision at stage t affects those at stage $t - 1$. Hence the influence of β^t on t -stage-decisions is spread to $t - 1$ -stage-decisions, as shown in Figure 4.7.

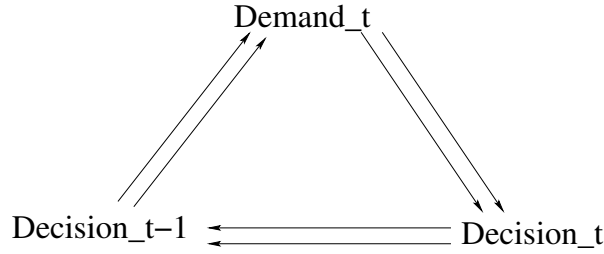


Figure 4.7: Influence Chart

4.6 Implementations of the Models

All models discussed in this chapter are written in AMPL [44] and solved by CPLEX, including the linearised form of model (4.12). Consider a 3-stage-problem, suppose each random variable has 2 possible values and there are 4 random variable sets, i.e., demand, capacities, initial capability and operation efficiency. Hence, there are 4369 nodes in total for this 3-stage-problem. At each node there are 7 integer decision variables and one continuous demand variable. Overall, there are 30583 decision variables in the model. In addition, 2 inventory constraints, one capacity constraint and 1 demand constraint at each node sum up to 17476 constraints. The data for the experiment can be found in the Appendix. They are generated according to our knowledge about banks without comprehensive scenario generation techniques.

This Aggregate Planning Model attempts to help decision makers to obtain the optimal decisions while satisfying the demand and controlling the risk. Our effort is focussed on risk management. As discussed above, qualified people armed with better knowledge and skills are considered to be safer for the company than non-qualified people. We test four different models: a basic model, a model with random capacity, a

model with BN revision, and a model with a demand function dependent on decisions. The basic model refers to the stochastic programming model assuming only randomness in the demand and no dependence between decisions and random variables, as given by (4.2) in Section 4.3. The model with random capacity corresponds to (4.3) - (4.5). The BN revision is governed by (4.11a), (4.11b) and (4.11c). The demand function depends on decisions through (4.12e), (4.12f) and (4.12g). The solution of the basic model excluding risk factors suggests employing more non-qualified people than the more elaborate models which aim for highest profit. The decisions provided by the basic model are questionable since they expose the company to the risk of failing to satisfy demand and operational errors. In the other models, we can see an average increase of employment of qualified people. Tables 4.2, 4.3 and 4.4 show the summary of results of each model. We report in them the numbers of qualified and non-qualified people to be employed as determined by an appropriate optimisation model. We solve a 3-stage problem but we are really concerned with the decisions at first stage and only these numbers are reported in the tables. In results presented in Tables 4.2 and 4.3, the risk is from random demand, random capacities and probability distributions, which are all principally resulting from market risks and consequently cannot be controlled. The analysis of results collected in Table 4.4, by considering the dependence between the decisions and variables, suggests that the controllable risks require more qualified people. By taking into account the operational risk in random capacity and the demand depending on decisions, and by constructing more reliable probability distribution, the model makes the decision to employ more qualified people than the basic model, so that the operation is more secure. Meanwhile, the cost, profit and loss are well balanced. In addition, the model with demand function depending on decisions reflects an average increase in demand of 5.48%.

Models	Basic Model	Random Capacity
Qualified People	2	4
Non-Qualified People	51	43

Table 4.2: Comparison between the basic model and the model with random capacity.

Models	Basic Model	BN Revision
Qualified People	2	6
Non-Qualified People	51	35

Table 4.3: Comparison between the basic model and the model with BN revision.

It is natural that the number of employees decreases when the corresponding cost (e.g. hiring, firing cost or salaries) increases. On the other hand, if people improve their skills, which means they can deal with more transactions or make fewer errors than before, they are more valuable to their employers. In our case study the parameter θ has

Models	Basic Model	Demand Function
Qualified People	2	9
Non-Qualified People	51	23

Table 4.4: Comparison between the basic model and the model with demand function depending on decisions.

more influence on decisions made than γ , i.e. a decrease in θ pushed down the number of qualified people employed. We have also looked at the evolution of employment over the planning period (3 stages) and we have observed that the number of non-qualified people does not change a lot. When demand varies from stage to stage, qualified people are more frequently fired or hired.

In the Appendix, we present some details of an approximated solution associated with the nonlinear model described by model (4.12).

4.7 The Pricing of Operational Risk

In the case of a convex nonlinear programming problem with equality and inequality constraints, the dual prices correspond to the Lagrange multipliers. Under certain condition, their interpretation is similar to that of the shadow prices in linear programming as explained in Chapter 2. For an additional unit of the right hand side parameter of a given constraint, the associated Lagrange multiplier indicates by how many units the objective function will vary, while other quantities remain the same.

In our context, the dual price of the constraint describing the availability of qualified workers gives the value of an additional hour of expertise provided by a qualified employee. In terms of risk management, we obtain pricing information to set up a kind of “strategic reserve” (terms from the military science designing a supplementary force available and ready to act in the case of urgent need).

Nowadays, in business this notion of strategic reserve for dealing with operational risk is not generally accepted. Generating a significant cost to hire expertise just to be able to solve difficult operations problems in case they might arise is not considered to be viable. However we believe that our model enables the risk budget planner to address the necessity to plan sufficient expertise in order to deal with unexpected operational problems. Moreover the dual solution represents a relevant way to quantify the expertise dedicated to risk management. The shadow price in the context of stochastic programming to produce a uniform CO_2 tax was first applied in a result analysis by Bahn et al. in [7].

The theory of shadow prices in linear programming was explained in Section 2.4.1. Before we apply this concept in a nonlinear model, it would be useful to carry out the sensitivity analysis of nonlinear programming in the following section.

4.7.1 The Sensitivity Analysis of Nonlinear Programming

Consider a nonlinear program of the form

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h_i(x) = 0, \quad i = 1, \dots, m, \\ & g_j(x) \leq 0, \quad j = 1, \dots, r, \end{aligned} \tag{4.14}$$

where f, h_i, g_j are continuously differentiable functions from \mathbb{R}^n to \mathbb{R} . A more succinct form could be

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h(x) = 0, \\ & g(x) \leq 0, \end{aligned} \tag{4.15}$$

where $h = (h_1, \dots, h_m) : \mathbb{R}^n \mapsto \mathbb{R}^m$, $g = (g_1, \dots, g_r) : \mathbb{R}^n \mapsto \mathbb{R}^r$. The corresponding Lagrangian function, with Lagrangian multipliers λ and μ , can be written as

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^r \mu_j(x) g_j(x).$$

A feasible point x is called *regular* if the constraint gradients $\nabla h_1(x), \dots, \nabla h_m(x), \nabla g_1(x), \dots, \nabla g_r(x)$ are linearly independent. If a local minimum of the problem \hat{x} is regular, and f, h_i, g_j are continuously differentiable, following [12], there exist unique Lagrange multiplier vectors $\hat{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_m)$, $\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_r)$, such that

$$\begin{aligned} \nabla_x L(\hat{x}, \hat{\lambda}, \hat{\mu}) &= 0, \\ \hat{\mu}_j &\geq 0, \quad j = 1, \dots, r, \\ \hat{\mu}_j &= 0, \quad \forall j \notin A(\hat{x}) = \{j | g_j(\hat{x}) = 0\}, \end{aligned}$$

where $A(\hat{x})$ is in fact the set of active inequality constraints at \hat{x} . In addition, if f, h_i, g_j are twice continuously differentiable, there holds

$$y' \nabla_{xx}^2 L(\hat{x}, \hat{\lambda}, \hat{\mu}) y \geq 0,$$

for all $y \in \mathbb{R}^n$ such that

$$\nabla h_i(\hat{x})' y = 0, \quad i = 1, \dots, m, \quad \nabla g_j(\hat{x})' y = 0, \quad j \in A(\hat{x}). \tag{4.16}$$

Conversely, we can have the sufficiency condition for (4.16) in Prop 6, called *Second Order Sufficiency Conditions*.

Proposition 6 ([12], Proposition 3.3.2, page 320). *Assume that f, h, g are twice*

continuously differentiable, and let $\hat{x} \in \mathbb{R}^n$, $\hat{\lambda} \in \mathbb{R}^m$, and $\hat{\mu} \in \mathbb{R}^r$ satisfy

$$\begin{aligned}\nabla_x L(\hat{x}, \hat{\lambda}, \hat{\mu}) &= 0, \\ h(\hat{x}) &= 0, \quad g(\hat{x}) \leq 0, \\ \hat{\mu}_j &\geq 0, \quad j = 1, \dots, r, \\ \hat{\mu}_j &= 0, \quad \forall j \notin A(\hat{x}) \\ y' \nabla_{xx}^2 L(\hat{x}, \hat{\lambda}, \hat{\mu}) y &\geq 0,\end{aligned}$$

for all $y \neq 0$ such that

$$\nabla h_i(\hat{x})' y = 0, \quad i = 1, \dots, m, \quad \nabla g_j(\hat{x})' y = 0, \quad j \in A(\hat{x}). \quad (4.17)$$

Assume also that

$$\hat{\mu}_j > 0, \quad \forall j \in A(\hat{x}).$$

Then \hat{x} is strict local minimum of f subject to $h(x) = 0$, $g(x) \leq 0$.

Similar to the linear program, Lagrange multipliers indicate the shadow prices of the constraints, e.g. resources, and provide sensitivity analysis as well. For nonlinear program, the sensitivity analysis is illustrated in the following proposition:

Proposition 7 ([12], Proposition 3.3.3, page 321). *Let \hat{x} , $\hat{\lambda}$, and $\hat{\mu}$ be a local minimum and Lagrange multipliers of (4.16), respectively, satisfying the second order sufficiency conditions of Prop 6, and assume that \hat{x} is a regular point. Consider the family of problems*

$$\begin{aligned}\min \quad & f(x) \\ \text{s.t.} \quad & h(x) = u, \\ & g(x) \leq v,\end{aligned} \quad (4.18)$$

parameterised by the vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^r$. Then there exists an open sphere \mathcal{S} centered at $(u, v) = (0, 0)$ such that for every $(u, v) \in \mathcal{S}$ there is an $x(u, v) \in \mathbb{R}^n$ and $\lambda(u, v) \in \mathbb{R}^m$, $\mu(u, v) \in \mathbb{R}^r$, which are a local minimum and associated Lagrange multiplier vectors of problem (4.18). Furthermore, $x(\cdot)$, $\lambda(\cdot)$, and $\mu(\cdot)$ are continuously differentiable in \mathcal{S} and we have $x(0, 0) = \hat{x}$, $\lambda(0, 0) = \hat{\lambda}$, $\mu(0, 0) = \hat{\mu}$. In addition, for all $(u, v) \in \mathcal{S}$, there holds

$$\begin{aligned}\nabla_u p(u, v) &= -\lambda(u, v), \\ \nabla_v p(u, v) &= -\mu(u, v),\end{aligned}$$

where $p(u, v)$ is the optimal cost parameterised by (u, v) ,

$$p(u, v) = f(x(u, v)).$$

We can see that one unit change of u_i (or v_j) contributes λ_i (or μ_j) units change of the objective function value in the opposite direction. The Lagrange multiplier λ_i associated with the equality constraint indeed measures the “force” of this equality constraint. For inequality constraint, if μ_j is a relatively large value compared to others, to improve the optimal objective value, the corresponding resource will be a higher priority to be increased than others.

Assume we relax the integer requirements of variables in model (4.2) in Section 4.3, model (4.2) combined with (4.3) and (4.5), model (4.10), and model (4.12) to be continuous. Then, all the models are linear except (4.12) which is nonlinear because of the dependence of random factor on decision variables in (4.12g). The general shadow prices theory can be applied to the first three models. For problem (4.12), the only nonlinear function appears in (4.12g) which involves an exponential function. It is easy to see that all feasible points of this problem are regular. And the objective function and constraint functions are twice continuously differentiable. Therefore, Proposition 7 can be applied to this model.

However, observe that in general Lagrange multipliers depend on the scaling of constraints. Hence, the same problem after scaling has a different Lagrange multiplier. In our case, all constraints have similar scaling and we can compare the associated Lagrange multipliers, or at the least we can use them to provide us with a qualitative insight.

The model is approximated to make it convex and smooth (though nonlinear). Keeping the continuity property enables us to produce shadow prices (Lagrange multipliers in the case of convex nonlinear models) which gives the implicit value of resources. In that case we obtain information related to value of expertise of qualified workers. To our knowledge this is the first time that shadow prices have been used to assess the cost of loss of workers in an operations risk management context.

4.7.2 Exploiting the Shadow Pricing Approach

In this section, we exploit the shadow prices of inventory constraints in four models, i.e. (4.2), (4.2) with (4.3) and (4.5), (4.10), and (4.12), in relaxed form that decision variables can take values from a continuous set. The inventory constraints (4.2b) and (4.2c), (4.10b), and (4.12b) express the balance between employees hired, fired and held at each stage for both qualified and non-qualified people. By keeping the original scaling of these constraints, we can compare the magnitude of the absolute values of Lagrange multipliers associated with these constraints and deduce from such a comparison which constraints are tight.

Tables 4.5, 4.6, 4.7 and 4.8 show the shadow prices of constraints in the four different models explained in Section 4.6. In each table, the absolute values of Lagrange multipliers associated with qualified people inventory constraints are in general greater than those of non-qualified people. In terms of risk management we interpret this fact as a warning that more attention should be paid to the availability of qualified workers

than that of non-qualified workers. For each node in every time period we identify the greatest danger to lose a qualified person. This can also be illustrated in the following shadow prices comparison tables. As we can see from Tables 4.6, 4.7 and 4.8 corresponding to the model with random capacity, the model with BN and the model with demand function depending on decisions, respectively, there is a general decrease in all absolute values of Lagrange multipliers in these three models, while the risk also shrinks.

Shadow Prices	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6
Inventory Constraint, Q People	7500	-2000	-7500	3000	-5000	2000
Inventory Constraint, NQ People	570	-1445	-1800	1650	-1075	1100
Capacity Constraints	30.5769	27.1154	14.5385	15	9.9808	10

Table 4.5: Shadow prices of inventory constraints and capacity constraints in the basic model.

Shadow Prices	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6
Qualified People	1645	810	-458	840	-146	560
Non-Qualified People	-81	-121	-75	372	36	248

Table 4.6: Shadow prices of inventory constraints in the model with random capacity.

Shadow Prices	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6
Qualified People	2028	-541	575	-230	383	-153
Non-Qualified People	136	-375	88	-109	62	-72

Table 4.7: Shadow prices of inventory constraints in the model with BN.

Shadow Prices	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6
Qualified People	7500	-2000	-3000	2000	-3000	2000
Non-Qualified People	430	-1304	-1650	1100	-1650	1100

Table 4.8: Shadow prices of inventory constraints in the model with demand function depending on decisions.

Chapter 5

Conclusions and Future Work

In this chapter, we summarise the work and results presented so far. The prospective research directions are discussed in the second part.

5.1 Research Outcomes

Risk management is essential to the financial market and a risk measure that can efficiently demonstrate the risks embedded in the financial products and instruments is always desirable. By the sources risks originates from, financial risk is divided into three categories: market risk, credit risk and operational risk. This research developed ways to measure and manage two of them, market risk and operational risk, under the discipline of stochastic programming.

Market risk dealt with in this work is that involved in ALM problems. In addition to the operational constraints, i.e. inventory and cash balance, ALM models require sophisticated risk control to ensure that liabilities are met. As a consequence, underfunding, which measures the amount of non-satisfied liabilities, is expected to be zero. Stochastic dominance as a standard of efficient risk control can manage the risk in ALM problems effectively and is consistent with utility theory. Furthermore, the concept of relaxed interval second-order stochastic dominance is developed and used to model chance constraints in linear form, which can manage underfunding in line with other stochastic dominance constraints. The object-oriented parallel solver OOPS [55, 57] can handle such problems efficiently in terms of both memory requirements and solution time.

In contrast to the simple structure of market risk, operational risk covers a wide range of risky events and has a variety of aspects. Consequently, complex modelling techniques are needed. The human factor in operational risk was analyzed and modelled following the methodology of the Aggregate Planning Model. Aggregate Planning Models are a category of mathematical programming model dealing with the basic productions or the operation management problems. In this thesis the focus is on the labour allocation management problem. By satisfying the demand constraints at each

stage, optimal staff allocation is determined while minimising costs (including salaries, hiring and firing costs) and losses, resulting from erroneous operations.

In the context of real enterprise risk management, decisions must be made that will affect future choices and outcomes. Hence when considering future events in business activity planning, it is pertinent to take into account uncertain parameters within the planning model. This is often done using a multistage stochastic programming model.

Although stochastic programming is a planning tool that simultaneously takes into account cause-and-effect relations and random variables, most applications in financial risks have been limited to the case where random variables are assumed to follow some theoretical probability distribution function. In order to add more relevance to the risk planning process of banking operations, we have combined the methodology of Bayesian networks with Aggregate Planning Models.

In general, the demand (a parameter of the model) is assumed to be independent of decisions. However, in reality, this is often not the case. If we consider for example the reputation of companies – the demand could be dependent upon the success of previous decisions; when customers receive products or services of high quality, they are more likely to continue the business and even increase its volume, which will increase the demand at next stage. Conversely, if customers are not satisfied with the service provided, they may change to other companies. This problem has been addressed in our stochastic Aggregate Planning Model by establishing a link between the random parameters and the decision variables. In particular, our model is in line with the COSO risk management philosophy which assumes that the quality of the Internal Control Systems affects the residual risks. Simply said, operations risks are controllable through good management, and it is an aspect that we take into account in our model for the first time.

This latter model results in a mixed integer problem that we have solved with CPLEX. Finally, we interpret interesting results obtained with this methodology that confirm that our modelling concept is relevant. Additionally, shadow prices of inventory constraints are used to price the risks of operations. Our model indicates at which period money should be set aside to be able to hire sufficient qualified workforce if needed.

This model of operational risk management is intended to support decision-making processes regarding employment strategies in order to manage operational risk from a human perspective. After studying the work efficiency and other kinds of skills of existing and potential employees, the management can input these coefficients into the model along with demand predictions. Then the model will generate an optimal strategy involving the proper workforce categories, repartitioning and minimising the risk of inadequate expertise. It must also be added that the tractability of the model presented means that it can be implemented and solved by most commercial optimisation codes.

5.2 Research Prospects

To further address the issue with work efficiency in operational risk management, we propose the application of learning curves to the model that describes the employees' learning process more precisely. In addition, the notion of service delay will also be worth incorporating into the model, which means we relax the assumption that there is enough temporary employees as back-up.

A significant research step will be the integration of the risk management, namely enclosing market risk, credit risk and operational risk management within a single model. This will require much more work, but the appealing ultimate objective of developing an integrated risk management model justifies such effort. Such an integrated system of risk management will facilitate risk management for industries and will help businesses better control their risks.

Bibliography

- [1] W. J. ABERNATHY, N. BALOFF, J. C. HERSHEY, AND S. WANDEL, *A three-stage manpower planning and scheduling model-a service-sector example*, Operations Research, 21 (1973), pp. 693–711.
- [2] S. AHMED, A. J. KING, AND G. PARIJA, *A multi-stage stochastic integer programming approach for capacity expansion under uncertainty*, Journal of Global Optimization, 26 (2003), pp. 3–24.
- [3] A. ALONSO-AYUSO, L. F. ESCUDERO, AND M. T. ORTUÑO, *Modelling production planning and scheduling under uncertainty*, in Applications of Stochastic Programming, S. W. Wallace and W. T. Ziemba, eds., SIAM Series on Optimization, 2002, pp. 217–252.
- [4] F. ANDERSSON, H. MAUSSER, D. ROSEN, AND S. URYASEV, *Credit risk optimization with conditional Value-at-Risk criterion*, Math. Program., 89 (2001), pp. 273–291.
- [5] Z. ARTSTEIN, *Gains and costs of information in stochastic programming*, Annals of Operation Research, 85 (1999), pp. 129–152.
- [6] P. ARTZNER, F. DELBAEN, J.-M. EBER, AND D. HEATH, *Coherent measures of risk*, Mathematical Finance, 9 (1999), pp. 203–228.
- [7] O. BAHN, E. FRAGNIÈRE, AND S. KYPREOS, *Swiss energy taxation options to curb CO₂ emissions*, European Environment, (1998), pp. 94–101.
- [8] W. M. BASSIN, *A Bayesian optimal overhaul interval model for the Weibull restoration process case*, Journal of the American Statistical Association, 68 (1973), pp. 575–578.
- [9] J. F. BENDERS, *Partitioning procedures for solving mixed-variables programming problems*, Numerische Mathematik, 4 (1962), pp. 238–252.
- [10] J. M. BERNARDO AND A. F. SMITH, *Bayesian Theory*, John Wiley and Sons, 2000.
- [11] D. BERNOULLI, *Exposition of a new theory on the measurement of risk*, Econometrica, 22 (1954), pp. 23–36.

- [12] D. P. BERTSEKAS, *Nonlinear Programming*, Athena Scientific, 2003.
- [13] J. R. BIRGE AND F. LOUVEAUX, *Introduction to Stochastic Programming*, Springer-Verlag, 1997.
- [14] E. BOGENTOFT, H. E. ROMELJN, AND S. URYASEV, *Asset/liability management for pension funds using CVaR constraints*, The Journal of Risk Finance, 3 (2001), pp. 57–71.
- [15] M. CARIO AND B. NELSON, *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix*, Northwestern University, IEMS Technical Report, (1997).
- [16] D. R. CARIÑO, T. KENT, D. H. MYERS, C. STACY, M. SYLVANUS, A. L. TURNER, K. WATANABE, AND W. T. ZIEMBA, *The Russell-Yasuda Kasai model: An asset/liability model for a Japanese insurance company using multistage stochastic programming*, Interfaces, 24 (1994), pp. 29–49.
- [17] D. R. CARIÑO AND W. T. ZIEMBA, *Formulation of the Russell-Yasuda Kasai financial model*, Operations Research, 46 (1998), pp. 433–449.
- [18] A. CHARNES AND S. THORE, *Planning for liquidity in financial institutions: The chance-constrained method*, Journal of Finance, 21 (1966), pp. 649–674.
- [19] T.-M. CHEN AND E. POPOVA, *Bayesian maintenance policies during a warranty period*, Communications in Statistics-Stochastic Models, 16 (2000), pp. 121–142.
- [20] A. J. CONEJO, F. J. NOGALES, J. M. ARROYO, AND R. GARCIA-BERTRAND, *Risk-constrained self-scheduling of a thermal power producer*, IEEE Transactions on Power Systems, 19 (2004), pp. 1569–1574.
- [21] G. CONSIGLI AND M. A. H. DEMPSTER, *Dynamic stochastic programming for asset-liability management*, Annals of Operation Research, 81 (1998), pp. 131–161.
- [22] A. CONSIGLIO AND D. D. GIOVANNI, *Evaluation of insurance products with guarantee in incomplete markets*, tech. report, University of Palermo, 2006.
- [23] —, *Pricing the option to surrender in incomplete markets*, tech. report, University of Palermo, 2007.
- [24] C. CORNALBA AND P. GIUDICI, *Statistical models for operational risk management*, Physica A, 338 (2004), pp. 166–172.
- [25] M. G. CRUZ, *Modeling, Measuring and Hedging Operational Risk*, John Wiley and Sons, 2002.

- [26] H. G. DAELLENBACH AND S. H. ARCHER, *The optimal bank liquidity: A multi-period stochastic model*, The Journal of Financial and Quantitative Analysis, 4 (1969), pp. 329–343.
- [27] J. DEBELY, M. DUBOSSON, AND E. FRAGNIÈRE, *The travel agent: Delivering more value by becoming an operational risk manager*, in 9th International Research Seminar in Service Management, 2006, pp. 178–203.
- [28] D. DENTCHEVA, R. HENRION, AND A. RUSZCZYŃSKI, *Stability and sensitivity of optimization problems with first order stochastic dominance constraints*, SIAM Journal of Optimization, 18 (2007), pp. 322–337.
- [29] D. DENTCHEVA AND W. RÖMISCH, *Optimal power generation under uncertainty via stochastic programming*, in Stochastic Programming Methods and Technical Applications, vol. 458, Springer, Berlin, 1998, pp. 22–56.
- [30] D. DENTCHEVA AND A. RUSZCZYŃSKI, *Optimization with stochastic dominance constraints*, SIAM Journal of Optimization, 14 (2003), pp. 548–566.
- [31] ———, *Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints*, Mathematical Programming, 99 (2004), pp. 329–350.
- [32] ———, *Portfolio optimization with stochastic dominance constraints*, Journal of Banking Finance, 30 (2006), pp. 433–451.
- [33] ———, *Optimization with multivariate stochastic dominance constraints*, Mathematical Programming, (2009), pp. 111–127.
- [34] M. DUBOSSON, E. FRAGNIÈRE, AND B. MILLIET, *A control system designed to address the intangible nature of service risks*, in IEEE International Conference on Service Operations and Logistics, and Informatics, 2006, pp. 90–95.
- [35] J. DUPAČOVÁ, N. GRÖWE-KUSKA, AND W. RÖMISCH, *Scenario reduction in stochastic programming*, Mathematical Programming, 95 (2003), pp. 493–511.
- [36] J. DUPAČOVÁ, *Optimization under exogenous and endogenous uncertainty*, in MME06, L. Lukáš, ed., University of West Bohemia in Pilsen, 2006.
- [37] J. DUPAČOVÁ, J. HURT, AND J. ŠTĚPÁN, *Stochastic Modeling in Economics and Finance*, Kluwer Academic Publishers, 2002.
- [38] B. P. DZIELINSKI AND R. E. GOMORY, *Optimal programming of lot sized, inventory and labour allocations*, Management Science, 11 (1965), pp. 874–890.
- [39] R. J. EBERT, *Aggregating planning with learning curve productivity*, Management Science, 23 (1976), pp. 171–182.

- [40] G. EPPEN, R. MARTIN, AND L. SCHRAGE, *A scenario approach to capacity planning*, Operations Research, 37 (1989), pp. 517–527.
- [41] C. I. FÁBIÁN, *Handling CVaR objectives and constraints in two-stage stochastic models*, Computational Management Science, 4 (2007), pp. 313–353.
- [42] C. I. FÁBIÁN, G. MITRA, AND D. ROMAN, *Processing second-order stochastic dominance models using cutting-plane representations*. Stochastic Programming E-Print Series, 10-2008.
- [43] P. C. FISHBURN, *Decision and Value Theory*, Wiley, 1964.
- [44] R. FOURER, D. M. GAY, AND B. W. KERNIGHAN, *AMPL: A Modeling Language for Mathematical Programming*, Thomson/Brooks/Cole, 2003.
- [45] E. FRAGNIÈRE AND J. GONDZIO, *Stochastic programming from modelling languages*, in Applications of Stochastic Programming, S. W. Wallace and W. T. Ziemba, eds., SIAM Series on Optimization, 2002, pp. 1–21.
- [46] E. FRAGNIÈRE, J. GONDZIO, AND X. YANG, *Operations risk management by optimally planning the qualified workforce capacity*, European Journal of Operational Research, 202 (2010), pp. 518 – 527.
- [47] E. FRAGNIÈRE AND G. SULLIVAN, *Risk Management-Safeguarding Company Assets*, Thomson, 2007.
- [48] J. GIERA, *BPO in banking: Outsourcing the back office*. Forrester Research, Nov 2004.
- [49] J. C. GITTINS AND K. D. GLAZEBROOK, *On Bayesian models in stochastic scheduling*, Journal of Applied Probability, 14 (1977), pp. 556–565.
- [50] V. GOEL AND I. E. GROSSMANN, *A class of stochastic programs with decision dependent uncertainty*, Mathematical Programming, 108 (2006), pp. 355–394.
- [51] R. GOLLMER, U. GOTZES, F. NEISE, AND R. SCHULTZ, *Risk modeling via stochastic dominance in power systems with dispersed generation*, tech. report, Department of Mathematics, University of Duisburg-Essen, 2007.
- [52] R. GOLLMER, U. GOTZES, AND R. SCHULTZ, *Second-order stochastic dominance constraints induced by mixed-integer linear recourse*, tech. report, Department of Mathematics, University of Duisburg-Essen, 2007.
- [53] R. GOLLMER, F. NEISE, AND R. SCHULTZ, *Stochastic programs with first-order dominance constraints induced by mixed-integer linear recourse*, SIAM Journal on Optimization, 19 (2008), pp. 552–571.

- [54] B. GOLUBA, M. HOLMERB, R. MCKENDALLC, L. POHLMANA, AND S. A. ZENIOS, *A stochastic programming model for money management*, European Journal of Operational Research, 85 (1995), pp. 282–296.
- [55] J. GONDZIO AND A. GROTHEY, *Parallel interior point solver for structured quadratic programs: Application to financial planning problems*, Annals of Operations Research, 152 (2007), pp. 319–339.
- [56] J. GONDZIO AND A. GROTHEY, *Solving nonlinear portfolio optimization problems with the primal-dual interior point method*, European Journal of Operational Research, 181 (2007), pp. 1019–1029.
- [57] J. GONDZIO AND A. GROTHEY, *Exploiting structure in parallel implementation of interior point methods for optimization*, Computational Management Science, 6 (2009), pp. 135–160.
- [58] S. C. GRAVES, *A review of production scheduling*, Operations Research, 29 (1981), pp. 646–675.
- [59] T. HAMADA AND K. D. GLAZEBROOK, *A Bayesian sequential single machine scheduling problem to minimize the expected weighted sum of flowtimes of jobs with exponential processing times*, Operations Research, 41 (1993), pp. 924–934.
- [60] G. HANOCH AND H. LEVY, *The efficiency analysis of choices involving risk*, The Review of Economic Studies, 36 (1969), pp. 335–346.
- [61] G. HANOCH AND H. LEVY, *The efficiency analysis of choices involving risk*, The Review of Economic Studies, (1969), pp. 335–346.
- [62] K. K. HAUGEN, *A stochastic dynamic programming model for scheduling of offshore petroleum fields with resource uncertainty*, European Journal of Operational Research, 88 (1996), pp. 88–100.
- [63] A. C. HAX AND H. C. MEAL, *Hierarchical integration of production planning and scheduling*, in Studies in Management Sciences, North Holland-American Elsevier, 1975.
- [64] M. HELLINGER, *Getting visibility into derivatives*, 2005. <http://www.FX-MM.com>.
- [65] F. S. HILLIER AND G. J. LIEBERMAN, *Introduction to Operations Research*, McGraw Hil, 2001.
- [66] ILOG, *Ilog cplex 10.0*. www.ilog.com.
- [67] F. V. JENSEN, *Bayesian Networks and Decision Graphs*, Springer, 2001.

- [68] T. W. JONBRÄTEN, R. J.-B. WETS, AND D. L. WOODRUFF, *A class of stochastic programs with decision dependent random elements*, Annals of Operations Research, 82 (1998), pp. 83–106.
- [69] P. JORION, *Financial Risk Manager Handbook*, Wiley, 2003.
- [70] J. B. KADANE, J. M. DICKEY, R. L. WINKLER, W. S. SMITH, AND S. C. PETERS, *Interactive elicitation of opinion for a normal linear model*, Journal of the American Statistical Association, 75 (1980), pp. 845–854.
- [71] P. KALL AND S. W. WALLACE, *Stochastic Programming*, John Wiley & Sons, 1994.
- [72] M. I. KAMIEN AND L. LI, *Subcontracting, coordination, flexibility, and production smoothing in aggregate planning*, Management Science, 36 (1990), pp. 1352–1363.
- [73] M. KAUT AND S. WALLACE, *Evaluation of scenario-generation methods for stochastic programming*, Stochastic Programming E-Print Series, 14 (2003).
- [74] D. KOS, *The changing nature of operational risk in foreign exchange*, Foreign Exchange Committee 2005 Annual Report, (2005).
- [75] L. J. KRAJEWSKI AND H. E. THOMPSON, *Efficient employment planning in public utilities*, The Bell Journal of Economics, 6 (1975), pp. 314–326.
- [76] P. KROKHMAL, J. PALMQUIST, AND S. URYASEV, *Portfolio optimization with conditional Value-at-Risk objective and constraints*, Journal of Risk, 4 (2001), pp. 43–68.
- [77] ———, *Portfolio optimization with conditional Value-at-Risk objective and constraints*, Journal of Risk, 4 (2001), pp. 43–68.
- [78] D. E. KROLL AND K. R. KUMAR, *The incorporation of learning in production planning models*, Annals of Operations Research, 17 (1989), pp. 291–304.
- [79] A. KÜNZI-BAY AND J. MAYER, *Computational aspects of minimizing conditional Value-at-Risk*, Computational Management Science, 3 (2006), pp. 3–27.
- [80] M. I. KUSY AND W. T. ZIEMBA, *A bank asset and liability management model*, Operations Research, 34 (1986), pp. 356–376.
- [81] P. L. BERNSTEIN AND A. DAMODARAN, *Investment Management*, John Wiley & Sons, Inc, 1998.
- [82] H. L. LEE, *Calculating risk capital: The OpRisk tree approach*, Global Association of Risk Professionals, (2006), pp. 10–15.

- [83] H. LEVY, *Stochastic dominance and expected utility: survey and analysis*, Management Science, 38 (1992), pp. 555–593.
- [84] H. LEVY AND M. SARNAT, *Portfolio selection and investors' utility: a graphical analysis*, Applied Economics, 2 (1970), pp. 113–119.
- [85] S. A. MALCOLM AND S. A. ZENIOS, *Robust optimization for power systems capacity expansion under uncertainty*, The Journal of the Operational Research Society, 45 (1994), pp. 1040–1049.
- [86] H. M. MARKOWITZ, *Portfolio Selection, Efficient Diversification of Investments*, John Wiley & Sons, 1959.
- [87] BASEL COMMITTEE ON BANKING SUPERVISION, *Operational risk, consultative paper 2*, 2001. Bank for International Settlements, Basel.
- [88] ———, *Basel 2: International convergence of capital measurement and capital standards, a revised framework*, 2004. Bank for International Settlements, Basel.
- [89] THE BOSTON CONSULTING GROUP, *What's next for India: Beyond the back of-fice*, 2005. Special Report.
- [90] P. MCCONNELL AND K. BLACKER, *An Approach to Modelling Operational Risk in Banks*, Henley Management College, 1999.
- [91] D. P. MORTON AND E. POPOVA, *A Bayesian stochastic programming approach to an employee scheduling problem*, IIE Transactions, 36 (2004), pp. 155–167.
- [92] J. M. MULVEY, *Applying optimization technology to portfolio management*, The Journal of Portfolio Management, (2004), pp. 162–168.
- [93] J. M. MULVEY, D. P. ROSENBAUM, AND B. SHETTY, *Strategic financial risk management and operations research*, European Journal of Operational Research, 97 (1997), pp. 1–16.
- [94] F. H. MURPHY AND Y. SMEERS, *Generation capacity expansion in imperfectly competitive restructured electricity markets*, Operations Research, 53 (2005), pp. 646–661.
- [95] M. P. NOWAK AND W. RÖMISCH, *Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty*, Annals of Operations Research, 100 (2000), pp. 251–272.
- [96] N. NOYAN, G. RUDOLF, AND A. RUSZCZYŃSKI, *Relaxations of linear programming problems with first order stochastic dominance constraints*, Operations Research Letters, 34 (2006), pp. 653–659.

- [97] N. NOYAN AND A. RUSZCZYŃSKI, *Valid inequalities and restrictions for stochastic programming problems with first order stochastic dominance constraints*, Tech. Report RRR 13-2006, Rutgers Center for Operations Research, 2006.
- [98] W. OGRYCAK AND A. RUSZCZYŃSKI, *From stochastic dominance to mean-risk models: Semideviations as risk measures*, European Journal of Operations Research, 116 (1999), pp. 33–50.
- [99] ———, *Dual stochastic dominance and related mean-risk models*, SIAM Journal of Optimization, 13 (2002), pp. 60–78.
- [100] H. H. PANJER, *Operational Risk : Modelling Analytics*, Wiley, 2006.
- [101] A. PARASURAMAN, V. A. AEITHAML, AND L. L. BERRY, *A conceptual model of service quality and its implications for future research*, Journal of Marketing, 49 (1985), pp. 41–50.
- [102] A. PASCUAL-LEONE, J. GRAFMAN, AND M. HALLETT, *Modulation of cortical motor output maps during development of implicit and explicit knowledge*, Science, 263 (1994), pp. 1287–1289.
- [103] G. PFLUG, *Online optimization of simulated Markov processes*, Mathematics of Operations Research, 15 (1990), pp. 381–395.
- [104] ———, *Some remarks on the Value-at-Risk and the conditional Value-at-Risk*, in Probabilistic Constrained Optimization: Methodology and Applications, S. Uryasev, ed., Kluwer Academic Publishers, 2000.
- [105] E. POPOVA AND D. MORTON, *Adaptive stochastic manpower scheduling*, in 1998 Winter Simulation Conference, 1998.
- [106] A. PRÉKOPA, *Stochastic Programming*, Kluwer Academic Publishers, 1995.
- [107] D. H. PYLE, *On the theory of financial intermediation*, Journal of Finance, 26 (1971), pp. 737–747.
- [108] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Scenarios and policy aggregation in optimization under uncertainty*, Mathematics of Operations Research, 16 (1991), pp. 119–147.
- [109] R. T. ROCKFELLAR AND S. URYASEV, *Optimization of conditional Value-at-Risk*, The Journal of Risk, 2 (2000), pp. 21–41.
- [110] D. ROMAN, K. DARBY-DOWMAN, AND G. MITRA, *Portfolio construction based on stochastic dominance and target return distributions*, Mathematical Programming, 108 (2006), pp. 541–569.

- [111] M. ROTHSCILD AND J. E. STIGLITZ, *Increasing risk: I. a definition*, Journal of Economic Theory, 2 (1970), pp. 225–243.
- [112] A. RUSZCZYŃSKI AND A. SHAPIRO, *Optimization of convex risk functions*, Mathematics of Operation Research, 31 (2006), pp. 433–452.
- [113] A. RUSZCZYŃSKI AND R. J. VANDERBEI, *Frontiers of stochastic nondominated portfolios*, Econometrica, 71 (2003), pp. 1287–1297.
- [114] G. H. SAAD, *Hierarchical production-planning systems: Extensions and modifications*, Journal of the Operational Research Society, 41 (1990), pp. 609–624.
- [115] T. SANTOSO, S. AHMED, M. GOETSCHALCKX, AND A. SHAPIRO, *A stochastic programming approach for supply chain network design under uncertainty*, European Journal of Operational Research, (2005), pp. 96–115.
- [116] S. SEN, R. D. DOVERSPIKE, AND S. COSARES, *Network planning with random demand*, Telecommunication Systems, (1994), pp. 11–30.
- [117] J. SMITH, *Moment methods for decision analysis*, Management science, 39 (1993), pp. 340–358.
- [118] R. M. SOLAND, *Bayesian analysis of the Weibull process with unknown scale and shape parameters*, IEEE Transactions on Reliability, (1969), pp. 181–184.
- [119] S. TAKRITI, J. R. BIRGE, AND E. LONG, *A stochastic model for the unit commitment problem*, IEEE Transaction on Power System, 11 (1996), pp. 1497–1508.
- [120] S. TAKRITI, B. KRASENBRINK, AND L. S.-Y. WU, *Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem*, Operations Research, 48 (2000), pp. 268–280.
- [121] L. TEFATSION, *Stochastic dominance and the maximization of expected utility*, The Review of Economic Studies, 43 (1976), pp. 301–315.
- [122] S. URYASEV, *Conditional Value-at-Risk: Optimization algorithms and applications*, Financial Engineering News, 14 (2000), pp. 1–5.
- [123] ———, *Introduction to the theory of probabilistic functions and percentiles (Value-at-Risk)*, tech. report, University of Florida, 2000.
- [124] H. P. WILLIAMS, *Model Building in Mathematical Programming*, Wiley, 1978.
- [125] W. L. WINSTON, *Introduction to Mathematical Programming*, Duxbury Press, 1995.
- [126] C. R. WOLF, *A model for selecting commercial bank government security portfolios*, The Review of Economics and Statistics, 51 (1969), pp. 40–52.

- [127] S. J. WRIGHT, *Primal-Dual Interior-Point Method*, SIAM, 1997.
- [128] H. Y. YAMIN, *Review on methods of generation scheduling in electric power systems*, Electric Power Systems Research, 69 (2004), pp. 227–248.
- [129] X. YANG, J. GONDZIO, AND A. GROTHEY, *Asset-liability management modelling with risk control by stochastic dominance*, Tech. Report ERGO-09-002, School of Mathematics and Maxwell Institute for Mathematical Sciences University of Edinburgh, January 16 2009.
- [130] Y. YASUDA, *Application of Bayesian Inference to Operational Risk Management*, PhD thesis, University of Tsukuba, 2003.
- [131] A. ZELLNER, *Optimal information processing and Bayes's theorem*, The American Statistician, 42 (1988), pp. 278–280.
- [132] W. T. ZIEMBA AND J. M. MULVEY, *Worldwide Asset and Liability Modeling*, Publications of the Newton Institute, Cambridge University Press, Cambridge, 1998.

Appendix A

Parameters and results of Model (4.12) in Chapter 4

This Appendix shows the data set and solutions to the model given by Equations (4.12).

Parameters set:

$$\begin{aligned} wageQ &= 60000, \\ wageNQ &= 13500, \\ wageX &= 300, \\ FirecostQ &= 12500, \\ FirecostNQ &= 3000, \\ Hirecost &= 5000. \end{aligned}$$

There are two possible values of the work capacities, operation efficiency and initial capabilities for both qualified and non-qualified people. They are given in Table A.1.

Employees	Work Capacities		Operation Efficiency		Initial Capabilities	
Qualified People	5.0	2.5	0.001	0.0015	0.75	0.55
Non-Qualified People	1.25	0.625	0.002	0.0025	0.75	0.55
Temporary	3.0	—	—	—	—	—

Table A.1: Work ability parameters.

The results are as follows:

Employees	Qualified	Non-Qualified	Temporary
No. Held	14	3	0

Table A.2: Decisions at first stage: numbers of people held.

node j	node l	node v	node s	Qualified	Non-qualified	Temp
1	1	1	1	11	3	0
1	1	1	2	11	2	17
1	1	2	1	11	3	0
1	1	2	2	11	2	17
1	2	1	1	14	3	1805
1	2	1	2	14	3	1805
1	2	2	1	14	3	1805
1	2	2	2	14	3	1805
2	1	1	1	12	1	25
2	1	1	2	12	1	25
2	1	2	1	12	1	25
2	1	2	2	12	1	25
2	2	1	1	14	3	2138
2	2	1	2	14	3	2138
2	2	2	1	14	3	2138
2	2	2	2	14	2	2192

Table A.3: Decisions at second stage: numbers of people held. Node j corresponds to demand state, l is the work capability, v is the initial capability and s is the operation efficiency.

node j	node l	node v	node s	Qualified	Non-qualified	Temp	Demand
1	1	1	1	8	3	34	11474
1	2	1	1	14	4	11	18474
1	3	1	1	8	3	33	11474
1	4	1	1	14	4	11	18474
2	1	1	1	8	3	37	11483
2	2	1	1	14	1	0	15575
2	3	1	1	12	0	0	15574
2	4	1	1	17	4	15	21574
3	1	1	1	12	0	0	15575
3	2	1	1	18	1	0	21575
3	2	1	3	17	4	10	21561

Table A.4: Part results at third stage: numbers of people held. Node j corresponds to demand state, l is the work capability, v is the initial capability and s is the operation efficiency.